THE PROCEEDINGS OF THE PHYSICAL SOCIETY

Section B

VOL. 69, PART 1 1 January 1956 No. 433 B

EDITORIAL

Sections A and B of the Proceedings

Attention is drawn to a new assignment of papers between Sections A and B of the *Proceedings*. The new subdivision, which will take effect from January 1956, is as follows :

Section A
(Physics of Elementary Processes)

Electrodynamics, Quantum Theory and Theory of Fields.

Molecular and Atomic Structure.

Nuclear Physics.

High Energy Particles and Cosmic Rays.

Atomic and Nuclear Collisions.

Section B
(Physics of Collective Processes)

*Physics (including electron physics) of Solids and Fluids.

Thermodynamics and Statistical Mechanics.

Low Temperature Physics.

Aerodynamics, Hydrodynamics and Acoustics.

Optics and Electron Optics.

*Physics of Biological Processes.

*Astrophysics, Geophysics and Atmospheric Physics.

*Electric Discharges.

*In certain circumstances where the main emphasis of the paper is on elementary processes a paper on this subject would appear in Section A.

Contributions to the *Proceedings* are divided into three categories : papers, research notes and letters to the editor.

Papers should be as concise as possible, and the publication of very long papers is discouraged. The presentation of work in a numbered series of separate papers is also usually discouraged.

Research Notes are intended for the communication of results where the work reported does not need a full paper. They should not normally exceed 4 printed pages, including diagrams, the equivalent length being about 2000 words if unaccompanied by diagrams.

Letters to the Editor are intended as a means of publication of new work which is of immediate importance to other workers in the field. They should not normally exceed 2 printed pages, including diagrams.

Further details of instructions to authors can be obtained from the Secretary-Editor at the offices of the Society.

Precision Measurements of Ionization Coefficients in Uniform Static Fields

By R. W. CROMPTON †, J. DUTTON and S. C. HAYDON ‡
Department of Physics, University College of Swansea

Communicated by F. Llewellyn Jones; MS. received 28th July 1955

Abstract. Recent work has shown that the growth of ionization currents in air and nitrogen up to high values of the parameter *pd* (where *p* is the gas pressure and *d* the gap distance) follows the well-known generalized Townsend equation and that the observed breakdown potentials are in agreement with those calculated from the Townsend breakdown criterion. Having established the nature of the general mechanism the next stage is to determine the ionization coefficients α and ω/α as accurately as possible. The present paper gives an analysis of the errors in the coefficients arising from experimental errors in the measurement of the ionization currents, and on the basis of the analysis defines the experimental conditions necessary for the accurate determination of the coefficients at high values of *pd* in hydrogen. The apparatus designed to satisfy these conditions is described, and the values of the ionization coefficients, determined for a range of *pd* from about 200 to 900 mm Hg cm in hydrogen, are given. Investigation of the effect of the state of the cathode surface on the secondary ionization coefficient showed that even at the comparatively high value of $pd \sim 350$ mm Hg cm in hydrogen, the cathode played a significant role in the growth of ionization currents and thus in the setting of the breakdown criterion.

§ 1. Introduction

The elucidation of the mechanism of the electrical breakdown of gases in uniform fields at high values of the parameter *pd* (p = pressure, d = gap distance) is of intrinsic and technological importance, and has aroused considerable interest over a number of years (Llewellyn Jones 1953, Meek and Craggs 1953, Raether 1949, Loeb and Meek 1941). The essence of the problem is the explanation of why at a certain minimum critical potential V_s, called the breakdown or sparking potential, a gas ceases to be an insulator and becomes a good conductor in which, if circuit conditions permit, large currents can develop. For convenience, two separate stages of the problem may be considered. First, it is necessary to explain in terms of fundamental physical processes why the transition of the gas from an insulator to a conductor occurs at a given minimum critical voltage V_s. In other words it is necessary to elucidate what mechanism sets the ***breakdown criterion*** for static uniform fields. Secondly, it is necessary to consider how the ionization current in the gas increases to its final glow discharge or arc value when a voltage equal to or greater than that given by the breakdown criterion is applied and maintained for a certain interval. The breakdown criterion is determined entirely by the electrical properties of the gas and electrode surfaces, whereas the subsequent development of the ionization current to high values depends also on the circuit conditions. This paper is concerned entirely with the growth of ionization currents at voltages V less than V_s (the so-called pre-breakdown ionization currents) and with the mechanism which sets the breakdown criterion in static uniform fields.

---

† On leave from Adelaide University, South Australia, 1953-54.
‡ Now at University of New England, Armidale, New South Wales.

Experiment (Llewellyn Jones and Parker 1950, 1952, Dutton, Haydon and Llewellyn Jones 1952) has shown that in air and nitrogen the growth of pre-breakdown ionization currents at a constant value E/p (E = field strength) at high pd is given by the same Townsend equation as that found to hold at low pd, viz.

$$I = I_0 e^{\alpha d}/\{1-(\omega/\alpha)(e^{\alpha d}-1)\}, \qquad \ldots\ldots(1)$$

where I_0 is a constant externally generated cathode current, d the distance between electrodes, α the primary ionization coefficient and ω/α a generalized secondary ionization coefficient. This generalized secondary coefficient can be approximately expressed (Llewellyn Jones and Parker 1950, 1952) as the linear sum of coefficients representing separately the action of photons δ, positive ions γ or metastable atoms ϵ at the cathode, or of positive ions β or photons η (under certain conditions) in the gas; i.e.

$$\omega = \beta + \alpha\gamma + \delta + \epsilon + \eta.$$

It follows from equation (1) that the criterion which gives the value of the minimum potential for static spark breakdown is given when the denominator of equation (1) vanishes, i.e. when

$$1-(\omega/\alpha)(e^{\alpha d}-1)=0, \qquad \ldots\ldots(2)$$

and this is in accord with experiment. In these conditions there is a continuous development of ionization in a field which is undistorted by space charge at least up to the setting of the breakdown criterion. The general mechanism by which the breakdown criterion is set at the higher values of the parameter pd is the same as at the lower values of pd. In order to proceed with the interpretation of the ionization coefficients in terms of fundamental atomic data, or to help to determine which of the possible secondary ionization processes are predominant in any given case, it is now desirable to determine the values of the coefficients themselves as accurately as possible for given conditions which include the state of the electrode surface as well as that of the gas. Since the coefficients are continuous functions of the field, an accurate knowledge of the values of these coefficients is also of importance in the study of the time rate of the initial growth of ionization current under impulse conditions when the applied potential is greater than the static sparking potential (Llewellyn Jones 1955). It is the purpose of this paper to describe the conditions necessary for the accurate determination of the ionization coefficients α and ω/α at high values of pd in hydrogen, the apparatus designed to meet these conditions, and, finally, to give some of the results and data obtained.

Hydrogen was chosen in the present investigation for the following reasons: It is a gas readily obtained in a relatively pure state; a great deal is known already about its electrical properties from experiments at low pressures (Townsend and Hurst 1904, Ayres 1923, Hale 1938, 1939 a, b, Llewellyn Jones and Davies 1951 a, b, Rose 1954); it is of increasing importance in industrial applications (for which a knowledge of its electrical properties is required); and the cathode surface used in the experiments may be cleaned by a glow discharge in the gas which is being investigated.

§ 2. PREVIOUS DATA AND PRELIMINARY RESULTS

The first measurements of ionization coefficients in hydrogen were carried out by Townsend and Hurst (1904) who measured α and ω/α over the range of E/p from about 40 to 90 v cm^{-1} (mm Hg)$^{-1}$ corresponding to low values of pd_s from about 4 to 12 mm Hg cm and gas pressures of the order of 20 mm Hg. The next investigation was that undertaken by Ayres (1923) who measured α over a range of E/p from 5 to 3000 v cm^{-1} (mm Hg)$^{-1}$ and ω/α over the range of E/p from 30 to 3000 v cm^{-1} (mm Hg)$^{-1}$. The range of the experiments did not permit the measurement of ω/α in the interesting range of E/p less than 30 v cm^{-1} (mm Hg)$^{-1}$. A more recent investigation of ionization coefficients in hydrogen was carried out by Hale (1938, 1939 a, b), who obtained values of α for E/p from 22 to 917 v cm^{-1} (mm Hg)$^{-1}$ at pressures of the order of 20 mm Hg and values of ω/α for E/p in the range c. 30 to 917 v cm^{-1} (mm Hg)$^{-1}$ but insufficient data were given in his paper for finding ω/α when $E/p < 30$ v cm^{-1}(mm Hg)$^{-1}$. Thus the primary coefficient α has been measured at the lower gas pressures over a range of E/p from 5 to 3000 v cm^{-1} (mm Hg)$^{-1}$, but the values of the secondary ionization coefficient ω/α at values of E/p less than 30 v cm^{-1} (mm Hg)$^{-1}$ for values of pd_s greater than 100 mm Hg cm (especially at higher gas pressures) has never been fully investigated. The first purpose of the present investigation was therefore to establish whether it is possible to measure secondary ionization coefficients in hydrogen in the range of low values of E/p from 20 to 25 v cm^{-1} (mm Hg)$^{-1}$ (corresponding to a value of pd_s from about 200 to about 1000 mm Hg cm) especially at high gas pressures. In order to do this, and also to obtain approximate values of α, preliminary experiments were first carried out using the ionization chamber and measuring technique previously used for experiments with air (Llewellyn Jones and Parker 1950, 195[illegible] and nitrogen (Dutton, Haydon and Llewellyn Jones 1952).

A value of 23 v cm^{-1} (mm Hg)$^{-1}$ for E/p was first chosen [illegible] a pressure of 150 mm Hg. The hydrogen was obtained by passing B.O.C. tank hydrogen through a large palladium thimble and the usual liquid-air trap. The pre-breakdown ionization currents I were measured at various distances, and the results are shown in figure 1 in which the ionization current amplification I/I_0

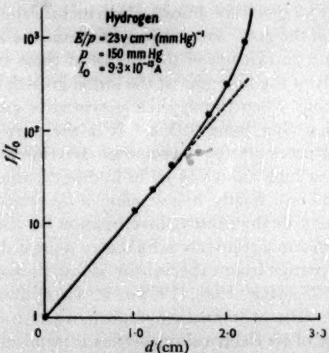


Figure 1. Curve of (log I/I_0, d) for hydrogen at high pressure: $E/p = 23$ v cm (mm Hg)$^{-1}$, $p = 150$ mm Hg, $I_0 = 9{\cdot}3 \times 10^{-13}$ A.

is plotted on a log scale against the distance d; I_0 was $9{\cdot}3 \times 10^{-13}$ A emitted from about 0·25 cm² of cathode. It can be seen that, as in the experiments with air and nitrogen, the $(\log I/I_0, d)$ curve is initially a straight line, but later curves upwards as the sparking distance d_s is approached. Results such as this are consistent with the operation of primary and secondary ionization processes in a uniform electric field. The value obtained for α is 2·77 and the value of ω/α is $1{\cdot}2 \times 10^{-3}$ (Crompton, Dutton and Haydon 1955).

The existence of a secondary ionization coefficient ω/α in this region having been established† it is now necessary to obtain values of α and ω/α as accurately as possible. In order to design an apparatus to measure these coefficients to any required degree of accuracy, it is first necessary to know how the errors in the values of the coefficients depend on the experimental errors in the measurement of the pre-breakdown ionization currents from which they are calculated. An analysis of this dependence will now be developed.

§ 3. ANALYSIS OF ERRORS

The coefficients are obtained from the Townsend relationship (equation (1)) in which there are in general three unknown quantities, viz. the coefficients α and ω/α, and (for reasons which will be discussed later) the value of I_0. In order to determine these three quantities, therefore, it is necessary to measure the ionization current corresponding to three suitable values of the electrode separation d. Even when considerable precautions are taken in experiments of this type, a difficulty arises in that the value of the initial photoelectric current I_0 may not remain absolutely constant during the course of the experiment: the measured values of I at the three different separations will then not correspond to an amplification of exactly the same initial current I_0, and errors will arise in the determination of α and ω/α. For accurate determination of the coefficients small variations in I_0 must be taken into consideration. It is therefore important to measure I_0, or some current which bears a constant relation to I_0, corresponding to each measured value of I. For reasons which appear later, it is the current $I_c = I_0/C$ (where C is a constant) which is measured experimentally in this investigation. Thus the three unknowns to be determined are now C, α and ω.

For the purpose of the analysis the following quantities are defined: I_1, I_2, I_3 are the ionization currents measured at the distances d_1, d_2, d_3 respectively, and I_{c1}, I_{c2}, I_{c3}, are the corresponding values of I_c. Also $R_1 = I_1/I_{c1}$, $R_2 = I_2/I_{c2}$, $R_3 = I_3/I_{c3}$, and $R_\alpha = R_2/R_1$, and $R = I_0/I_3 = CI_{c3}/I_3$.

Using the measured values of I and I_c the coefficients α and ω/α are calculated as follows.

Approximate values α_1 and C_1 of α and C are first calculated from the ratios R_1 and R_2 on the assumption that the contribution of the secondary effect ω/α to the currents may be neglected. Using these values of α_1 and C_1 an approximate value of ω/α is obtained from R_3. In practice d_1 may be chosen so small that the contribution from the secondary effect is in fact negligible. On the other hand $d_2 - d_1 = d$, say, must be large in order that errors in current and distance measurement should not themselves result in large errors in the calculated values of α.

---

† *Note added in proof:* Since this paper was written a letter by A. Wilkes, W. Hopwood and N. J. Peacock has been published in *Nature*, 1955, **176**, 837, giving measurements of I/I_0 for hydrogen up to a value of $pd_s \sim 2000$ mm Hg cm showing that the Townsend relationships (1) and (2) apply in this range.

Thus α_1 may differ from the true value of α because the contribution from ω/α to R_2 is not negligible. Let the true value be $\alpha=\alpha_1-\Delta\alpha$, then

$$e^{\Delta\alpha d}=1/\{1-(\omega/\alpha)(e^{\alpha d}-1)\}. \qquad \ldots\ldots(3)$$

The true value of the correction $\Delta\alpha$ may only be obtained if the true values of α and ω/α are known, but increasingly accurate values of α and ω/α may be obtained by using the method of successive approximations.

It is now possible to proceed to the analysis of errors in the coefficients arising from experimental errors in the measurement of currents.

The error in determination of α is

$$\frac{\Delta\alpha}{\alpha}=\frac{1}{\alpha d}\frac{\Delta R_\alpha}{R_\alpha}, \qquad \ldots\ldots(4)$$

and it is clear that d must be made large if the percentage error in α due to errors in the measurement of currents is to be small.

The error in determination of ω/α arises both directly through the influence of errors in current measurement on R_ω and indirectly through the influence of the errors on α. Considering these in turn it can be seen from the equation for ω/α which is

$$\omega/\alpha=\{1-R_\omega\exp(\alpha d_3)\}/\{\exp(\alpha d_3)-1\} \qquad \ldots\ldots(5)$$

that the error in ω/α arising from an error in R_ω is given by

$$\left[\frac{\Delta(\omega/\alpha)}{(\omega/\alpha)}\right]_\alpha=\left[\frac{-R_\omega\exp(\alpha d_3)}{1-R_\omega\exp(\alpha d_3)}\right]\frac{\Delta R_\omega}{R_\omega}. \qquad \ldots\ldots(6)$$

The error in ω/α arising from an error in α on the other hand, is given by

$$\left[\frac{\Delta(\omega/\alpha)}{(\omega/\alpha)}\right]_{R_\omega}=\left[\frac{\alpha d_3\exp(\alpha d_3)(R_\omega-1)}{\{1-R_\omega\exp(\alpha d_3)\}\{\exp(\alpha d_3)-1\}}\right]\frac{\Delta\alpha}{\alpha}. \qquad \ldots(7)$$

Now when d_3 approaches the sparking distance d_s, $R_\omega\ll 1$ and $\exp(\alpha d)\gg 1$ so that equations (6) and (7) reduce to

$$\left[\frac{\Delta(\omega/\alpha)}{(\omega/\alpha)}\right]_\alpha=\left[-\frac{R_\omega}{\omega/\alpha}\right]\frac{\Delta R_\omega}{R_\omega}=-X\frac{\Delta R_\omega}{R_\omega} \qquad \ldots\ldots(8)$$

and

$$\left[\frac{\Delta(\omega/\alpha)}{(\omega/\alpha)}\right]_{R_\omega}=\left[\frac{-\alpha d_3}{\{1-R_\omega\exp(\alpha d_3)\}}\right]\frac{\Delta\alpha}{\alpha}=-Y\frac{\Delta\alpha}{\alpha}. \qquad \ldots\ldots(9)$$

In order to apply the analysis to a particular case it is necessary to know the approximate values of α, ω/α and d_s for the gas to be studied. In the case of hydrogen the preliminary experiments indicated that at an E/p of 23 v cm^{-1}(mm Hg)$^{-1}$, $\alpha\sim 2{\cdot}7$, $\omega/\alpha\sim 1\times 10^{-3}$ and $d_s\sim 2{\cdot}5$ cm. Using these values, the way in which the quantities X and Y of equations (8) and (9) vary with distance is shown in figure 2. From this it can be seen that for distances d_3 within 10% of the sparking distance the error in ω/α, due to an error in the measurement of current, becomes much smaller than the error in ω/α due to the error in α. Thus, since in practice it is possible to make measurements for all distances d_3 up to and including a distance which is 5% less than the sparking distance, the principal cause of error in ω/α arises out of an inaccuracy in the calculated value of α. For this particular case when d_3 is taken as 2·4 cm it can be seen that $\Delta(\omega/\alpha)/(\omega/\alpha)\simeq -10\Delta\alpha/\alpha$. Thus to reduce the error in ω/α to 10% it is necessary to know α to within 1·0% and equation (2) shows immediately that if, for example, $d=1{\cdot}4$ cm, $\Delta\alpha/\alpha=\frac{1}{4}\Delta R_\alpha/R_\alpha$ so that R_α must be determined within 4%. Thus since R_α involves the measurement of four currents, each current must be measured to 1%. This then is the accuracy to be aimed at in designing a measuring apparatus to measure α to 1% and ω/α to 10%.

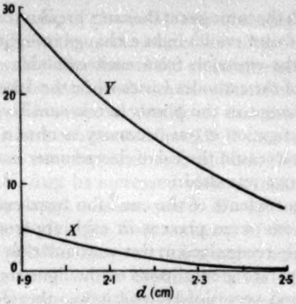


Figure 2. Graph of the factors X and Y (equations (8) and (9)) as functions of the distance.

§ 4. APPARATUS

The apparatus designed to obtain this degree of precision was essentially the same as that used in previous investigations in air (Llewellyn Jones and Parker 1950, 1952) and nitrogen (Dutton, Haydon and Llewellyn Jones 1952) except for important modifications in the current measuring technique. Since it is these modifications which have led to the high degree of accuracy now attainable in the measurement of α and ω/α the technique of current measurement will be given in detail.

4.1. *Method of Measuring Ionization Currents*

The method adopted was to measure, by means of an electrometer valve, the potential difference developed across a high resistance through which the ionization current flows. The advantage of this method is that it is possible to measure currents continuously and quickly. The relevance of these two considerations will be discussed later.

The circuit of the electrometer used was that described by Crompton and Sutton (1952). In order to measure currents in the range 10^{-13} to 10^{-8} A, a set of high stability resistors (10^{11} to 10^{7} ohms) was mounted on a high resistance switch so that any one could be connected between the grid and earth. The values of resistors with nominal values from 10^{11} to 10^{8} ohms were measured accurately by means of a Townsend induction balance method, and the resistor of nominal value 10^{7} ohms was measured by comparison with a standard megohm by means of a bridge method. For reasons discussed below a set of four high insulation Suflex condensers (50 to 50 000 pF) were also arranged so that any one could be connected between the grid of the electrometer and earth.

4.2. *Stability of the Initial Photoelectric Current*

Since it was required to measure the ratio of the gas-amplified current to the initial current from the cathode to within 2%, it is clearly necessary that the cathode emission should be stable. Previous experiments (Llewellyn Jones and Parker 1950, 1952, Dutton, Haydon and Llewellyn Jones 1952) have shown that there are fluctuations in cathode emission which may be due either to a variation of the photoelectric sensitivity of the surface or to a variation in the intensity of the irradiation. In order to investigate the relative importance of these two processes a control phototube was constructed having electrodes of similar material to those in the ionization chamber, and both the phototube and chamber were filled with the same gas at the same pressure. If both are illuminated with the same source of ultra-violet light a change in light intensity should result in similar changes in the emission from each cathode. If, on the other hand, the emission from one of the cathodes varies while the other remains constant then this must indicate a change in the photoelectric sensitivity of the first cathode. To carry out this investigation it was necessary to obtain readings of the current continuously and for this reason the valve electrometer in conjunction with a high resistance described above was used.

Simultaneous measurements of the emission from each cathode showed that both types of fluctuations were present in each apparatus. The variations in light intensity were easily recognized in this way, and they were reduced considerably by adjusting the operating conditions of the light source. The fluctuations which remained occurred very slowly, and it was therefore considered that the required precision could be obtained provided a measurement of the initial cathode emission was made prior to each measurement of the gas amplified ionization current.

The value of the initial emission I_0 is usually obtained by measuring the current in the gap when the applied voltage is sufficiently large for the current to saturate but insufficient to cause appreciable ionization. However, the (V, I) characteristic for a given gap separation in hydrogen showed no definite plateau, so that the value of the voltage (or $V/pd = E/p$) at which the current saturated could not be ascertained. It was therefore decided to obtain a measure of I_0 in these experiments by measuring the current I_c flowing when E/p was $5\,\mathrm{v\,cm^{-1}(mm\,Hg)^{-1}}$. Under these conditions there would be no appreciable ionization and I_c would be equal to some fraction of I_0.

In order to eliminate any effects due to the variation of photoelectric sensitivity of the cathode over its area the distances chosen were always integral multiples of a complete turn of the screw thread on which the cathode was mounted, so that the ultra-violet illumination always fell on the same part of the cathode (Dutton, Haydon and Llewellyn Jones 1952).

4.3. *Stability Requirements of the High Voltage Supply*

There are two factors which necessitate the use of high stability power supplies in these experiments. Firstly, the current is an exponential function of α, and α itself varies rapidly with the field, so that the current is highly dependent on the field, and thus on the voltage. Secondly, the ionization gap and the current measuring apparatus are two condensers in series across which the potential is applied, so that any fluctuations in the voltage across the system as a whole give rise to corresponding fluctuations on the measuring apparatus. The second consideration imposes the greater stringency on the voltage stability, so this will now be considered in more detail.

For a given fluctuation in the applied voltage the variation in voltage on the electrometer may be reduced by increasing the capacity to earth of the electrometer system. In the high resistance method, however, this capacity C_2 is limited by consideration of the time constant of the circuit. For example the highest value of resistance used in these experiments is 10^{11} ohms so that in order to obtain a reasonable value of the time constant for this resistor, say 50 seconds, the value of C_2 cannot be greater than 500 pF.

Assuming the capacity K_1 of the ionization chamber is 10 pF then to obtain fluctuations of the voltage on the electrometer less than 5×10^{-4} v (corresponding to $\pm 0{\cdot}5\%$ of the full scale deflection of the electrometer) ΔV must be less than 0·025 v. The currents to be measured with the resistance of 10^{11} ohms necessitate voltages up to 2500 v. This means, therefore, that the power supply must be stable to 10^{-3} %. When currents greater than 10^{-12} A are to be measured, however, resistances of lower values may be employed, and it is clear that in these cases the value of the condenser K_2 may be increased without making the value of the time constant inconveniently large. Thus when $R = 10^{10}$ ohms a capacity K_2 of 5000 pF may be employed giving a time constant of 50 seconds. Fluctuations in the high-voltage supply which are ten times larger may now be tolerated.

The voltage supply used in the experiments on air and nitrogen varied by about 5×10^{-2} %. Owing to fluctuations in the power mains in the immediate post-war period when this supply was first constructed, it was fed from a d.c. motor alternator run off large storage batteries. Improved regulation in the mains supply has now made it possible to replace this alternator by an a.c. constant voltage transformer. This modification, together with an alteration in the means of voltage control from a Variac to a step-by-step transformer in conjunction with a resistance box, has improved the stability of the supply by a factor of five, i.e. to 10^{-2} %. This was sufficient to enable currents 10^{-12} A to be measured to the required accuracy, but in order to measure currents less than this value a new power supply was constructed. This supply consisted of a simple half-wave rectifier with resistance–capacity smoothing which was supplied from a constant voltage transformer and which fed a resistance in series with twenty 150 B2 high stability voltage stabilizer tubes. The output could be varied in steps of 150 v up to 3 kv by means of a selecting switch, and a fine voltage control was obtained by a potentiometer connected across two of the stabilizer tubes. The short term stability obtained with this simple circuit was 1 part in 10^5.

This stability was still insufficient for the direct measurement of currents less than 10^{-12}A. These currents were therefore measured by a method similar to that used in the previous experiments on nitrogen. A dummy system was constructed, exactly similar electrically to the measuring system, and the sensitivity of the dummy electrometer adjusted to be the same as that of the measuring electrometer. Thus when a current in the ionization chamber was being measured as a deflection on the measuring electrometer, the dummy electrometer gave the deflection due to voltage fluctuations and hence the correction to be applied. The final procedure for the measurement of the ratio of I/I_0 at any distance was to set a voltage on the gap from the 0–3 kv voltage supply to give a value of E/p equal to 5 v cm^{-1} (mm Hg)$^{-1}$, and measure the current by the dummy system method. The voltage was then quickly changed to a pre-set value on the high voltage supply to give E/p in the range 20 to 25 v cm^{-1} (mm Hg)$^{-1}$, and the current measured directly.

§ 5. RESULTS

In addition to obtaining precise determinations of the primary and secondary ionization coefficients at low values of E/p, experiments were also carried out to ascertain whether secondary ionization processes at the cathode were playing a significant role in the growth of ionization current, and hence in breakdown, at these high values of pd_s. The method used was to investigate the effect on the coefficient ω/α of a change in the cathode surface, by measuring the growth of ionization currents in the same gas but with different cathode surfaces.

It was clearly essential in this work to obtain gas samples of consistent (although not necessarily absolute) purity. It was found that by repeatedly pumping the ionization chamber to a pressure of the order of 10^{-4} mm Hg, filling with hydrogen which had been passed through the purifying train of heated palladium and liquid air, allowing the gas to stand in the chamber some time, and then pumping out again, a consistent sample of hydrogen could eventually be obtained. Samples were judged to be of consistent purity when successive gas fillings gave values of α/p which agreed within the experimental error. When such consistent samples had been obtained, accurate measurements of ionization currents were made for $E/p=23$ v cm^{-1} (mm Hg)$^{-1}$ and $p=150$ mm Hg with a silver-plated cathode. From these results the coefficients α and ω/α were calculated by the method outlined above. A typical set of results is given in table 1 to show that an error less than 2% in the ratios of ionization current to initial current and of less than 1% in α had in fact been achieved. The cathode was subsequently treated with successive glow discharges at a current of the order of 150 mA in hydrogen at a pressure of a few mm Hg and the growth of ionization currents measured at intervals during the treatment. The results, which are given in table 2, show

Table 1. Set of Results for $E/p=23$ v cm^{-1} (mm Hg)$^{-1}$, $p=150$ mm Hg
Silver Cathode

d(cm)	$I_c \times 10^{13}$(A)	$I \times 10^{13}$(A)	I/I_c	Mean I/I_c
0·300	5·90	21·5	3·64	3·625
	5·95	21·5	3·61	
1·5699	6·15	660	107	105·7
	6·70	705	105	
	5·55	585	105	
	6·68	710	106	
1·6677	6·42	920	143	142
	6·40	905	141	
	6·25	885	142	
	6·85	970	142	
2·3515	6·60	10800	1636	1650
	6·4	10650	1664	
2·4492	7·47	22200	2972	2970
	6·87	20400	2969	

The mean values of α and ω/α calculated for these results are $\alpha=2{\cdot}61$, $\omega/\alpha=1{\cdot}15\times10^{-3}$.

Table 2. Values of α and ω/α at $E/p=23$ v cm^{-1} (mm Hg)$^{-1}$, $p=150$ mm Hg for a Silver Cathode subjected to a Glow Discharge

Treatment of Cathode	α	$\omega/\alpha \times 10^3$
Polished before assembly; one spark and a large number of ionization current measurements	2·60	1·05
One further spark and more ionization current measurements	2·61	1·15
Low pressure glow discharge in H_2 at about 150 mA for 20 min	2·61	1·29
Low pressure glow discharge in H_2 at about 175 mA for 200 min	2·60	1·20

that consistent values were obtained for α/p throughout the measurements, indicating that the gas samples were of consistent purity; thus the variations observed in ω/α, although small and showing no consistent trend with cathode treatment, were probably due to changes in the cathode surface. That such surface changes were occurring was confirmed by variation in the photoelectric sensitivity and the formation of surface layers during the treatment.

Measurements were next made with the silver-plated cathode replaced by a nickel cathode. The coefficients were measured for the initial samples of hydrogen; several (~20) sparks were then obtained at a voltage of about 8 kv and the coefficients remeasured. The apparatus was then dismantled, the cathode repolished, and the coefficients measured again in the first sample of hydrogen admitted after reassembly. The results, which are given in table 3, show that there is a strong indication that the secondary ionization coefficient had been altered by the effect of the sparks on the cathode surface. This is in agreement with the observation that the breakdown potential rose as the sparking proceeded. Measurements of the photoelectric current I_c can be used as an indication of any change in the photoelectric sensitivity of the cathode for radiation from the mercury arc, and thus comparison of I_c with the corresponding values of ω/α is of interest. In the above experiments the effect of the sparks was to reduce I_c from $\sim 2\times10^{-12}$ A to 2×10^{-13} A when ω/α was reduced from $2{\cdot}05\times10^{-3}$ to $1{\cdot}42\times10^{-3}$, indicating that the sensitivity of the surface is changed much more with respect to the external ultra-violet radiation than to the processes giving rise to the secondary ionization.

Table 3. Values of α and ω/α at $E/p=23$ v cm^{-1}(mm Hg)$^{-1}$, $p=150$ mm Hg for a Nickel Cathode subjected to High Pressure Sparking

State of Cathode	α	$\omega/\alpha\times10^3$
Cathode polished before assembly	2·57	2·08
Cathode subjected to about 20 sparks at ~8 kv	2·53	1·42
Cathode repolished	2·56	2·05

After repeated fillings with hydrogen, successive samples were found to give the same values of α/p as were obtained previously in the experiments with a silver cathode. The coefficients α and ω/α were then measured for a range of E/p from 20 to 25 v cm^{-1}(mm Hg)$^{-1}$. Since the purpose of the investigation was to obtain numerical data as exact as possible the actual values obtained are given in table 4 rather than in the form of a curve. It will be noted that with $E/p=23$ v cm^{-1}(mm Hg)$^{-1}$ the same value was obtained for α/p as when the silver cathode was used, but that the value of ω/α was about 50% greater than for the silver cathode. This result is significant since the difference in ω/α is well outside the experimental error, and indicates that the nature of the cathode had a definite

Table 4. Values of α and ω/α for E/p from 20 to 25 v cm^{-1}(mm Hg)$^{-1}$ for a Nickel Cathode

E/p (v cm^{-1} (mm Hg)$^{-1}$)	20	21	22	23	24	25
p(mm Hg at 20°c)	360	220	150	150	100	100
$\alpha/p\times10^2$ (cm^{-1} (mm Hg)$^{-1}$)	0·75	1·04	1·33	1·74	2·20	2·80
$\omega/\alpha\times10^3$	0·89	1·35	1·90	1·88	3·27	2·79
pd_s (calc.) (mm Hg cm)	934	636	470	361	260	210

influence on the value of ω/α. These results therefore show that the change in the nature of the base metal of the cathode from silver to nickel and the treatment of the nickel cathode with high pressure sparks had more effect on the secondary ionization process than did the glow discharge treatment given to the silver cathode.

Since the occurrence of a spark discharge in the ionization chamber can significantly change the cathode surface, no spark was allowed to occur during the set of results shown in table 4 until the measurements of α and ω/α had been completed. However, after this was done, a single but very careful determination of the static breakdown potential when E/p was 20 v cm^{-1} (mm Hg)$^{-1}$ was made by slowly increasing the separation d during measurement of a ($\log I, d$) curve, from which α and ω/α had already been found to be 2·71 and $0{\cdot}89 \times 10^{-3}$ respectively. Substitution of these values in the Townsend criterion (2) gave $d_s = 2{\cdot}596$ cm and $V_s = 18{\cdot}69$ kv. The experimentally measured value of V_s was within 0·1% of this value at $d = 2{\cdot}596$ cm.

It is of interest to note also that the results obtained for α/p in these samples of hydrogen, which were not necessarily of a high degree of purity, are in good agreement with the results recently obtained by Rose (1954, 1955) using extremely pure gas. The range of E/p was the same in both cases, but Rose used low gas pressures compared with those used in the present work. This agreement in values of α/p at the same E/p but different pressures is direct experimental evidence that the similarity principle $\alpha/p = f(E/p)$ holds for primary ionization over a wide range of pressures from a few mm Hg to 350 mm Hg in hydrogen. It follows that the mechanism of primary ionization is the same over this range of pressures and that the applied field is not appreciably distorted by the electron avalanche in this range.

§ 6. CONCLUSIONS

The following summarizes the conclusions drawn from these experiments:

1. The mechanism of the growth of pre-breakdown ionization currents in hydrogen up to a value of pd of the order of 1000 mm Hg cm is the same as that at the lower values of pd less than 50 mm Hg cm, being characterized by the operation of primary and secondary ionization processes in a uniform field.
2. Cathode secondary ionization processes play an important role in the growth of ionization currents, and thus in setting the breakdown criterion at high values of pd_s in uniform static fields in hydrogen.
3. Agreement between results for the coefficient α/p at low and high pressures gives direct experimental evidence for the similarity relationship $\alpha/p = f(E/p)$ in hydrogen.
4. A method has been developed which enables the ionization coefficients to be measured to the order of accuracy sought.

ACKNOWLEDGMENTS

We wish to acknowledge the support we have received from the Central Electricity Authority in this investigation and to thank them in particular for an appointment held by one of us (S.C.H.). We are grateful to Professor Llewellyn Jones, in whose laboratory this work was carried out, for his help and interest throughout the investigation. We also wish to thank Mr. D. K. Davies for help in taking measurements during the later stages of this work.

References

AYRES, T. L. R., 1923, *Phil. Mag.*, **45,** 353.

CROMPTON, R. W., DUTTON, J., and HAYDON, S. C., 1955, *International Symposium on Electrical Discharges, Delft*, April 1955. *Appl. Sci. Res.* B, **5,** 111.

CROMPTON, R. W., and SUTTON, D. J., 1952, *Proc. Roy. Soc.* A, **215,** 467.

DUTTON, J., HAYDON, S. C., and LLEWELLYN JONES, F., 1952, *Proc. Roy. Soc.* A, **213,** 203.

HALE, D. H., 1938, *Phys. Rev.* **54,** 241 ; 1939 a, *Ibid.*, **55,** 815; 1939 b, *Ibid.*, **56,** 1149.

LLEWELLYN JONES, F., 1953, *Rep. Prog. Phys.*, **16,** 216 (London : Physical Society); 1955, *Handbuch der Physik*, Vol. 22, Growth of Ionization and Breakdown (Heidelberg : Springer).

LLEWELLYN JONES, F., and DAVIES, D. E., 1951 a, *Proc. Phys. Soc.* B, **64,** 397; 1951 b, *Ibid.*, **64,** 519.

LLEWELLYN JONES, F., and PARKER, A. B., 1950, *Nature, Lond.*, **165,** 960; 1952, *Proc. Roy. Soc.* A, **213,** 185.

LOEB, L. B., and MEEK, J. M., 1941, *The Mechanism of the Electric Spark* (Stanford : University Press).

MEEK, J. M., and CRAGGS, J. D., 1953 *Electrical Breakdown of Gases* (Oxford : Clarendon Press).

RAETHER, H., 1949, *Ergebn. exakt. Naturw.*, **22,** 273.

ROSE, D. J., 1954, *Conference on Gaseous Electronics*, New York; see also 1955, *Nature, Lond.*, **175,** 154.

TOWNSEND, J. S., and HURST, H. E., 1904, *Phil. Mag.*, **8,** 738.

The Work Function and Patch Field of an Irregular Metal Surface

By M. J. MORANT † and H. HOUSE
Electrical Engineering Department, Queen Mary College, London

MS. received 16*th May* 1955, *and in final form* 14*th September* 1955

Abstract. In removing an electron from a metal, work has to be done against image-type forces. The classical image potential outside projections on an irregular surface is less than outside a plane surface but it is shown that the resulting lowering of the work function is entirely compensated by a patch field. Exact calculations cannot be made even for the single idealized surface projection for which image calculations are available. However, the axial patch potential for this case has been evaluated using an electrolytic tank and this shows that the form of the surface barrier is not affected by surface microgeometry.

---

§ 1. INTRODUCTION

It is well known that, due to the variations of the work function over the surface, the characteristics of electron emission from real metals in high accelerating fields are considerably different from those of ideal metals. Such variations are usually associated with adsorbed layers or the multicrystalline structure of the metal. As electron emission is very sensitive to small changes of work function, it is important to consider whether the microgeometry of the surface can influence the height or shape of the image barrier and so cause localized variations of the work function.

An electron at a distance x from a conductor induces a distributed charge of $+\mathbf{e}$ on the surface. If this is plane over an area large compared with x the attractive force between the electron and the induced charge is given by the classica image law

$$F_1 = \mathbf{e}^2/4x^2. \qquad \ldots\ldots(1)$$

This law obviously fails for small values of x and it has to be replaced, at distances below x_0, by quantum mechanical calculations of the interaction energy of the electron and the metal (Bardeen 1936 a, 1940, Sachs and Dexter 1950). Estimates of x_0 may also be obtained from the theory of the surface contribution to the work function. A calculation of this type for sodium has been given by Bardeen (1936 b) and slight improvements have been suggested by Juretschke (1953). It is reasonable to assume that the main features of the surface potential barrier of sodium are the same as for other metals for which detailed calculations cannot yet be made. Bardeen showed that the surface barrier was largely accounted for by the variation of the exchange-correlation potential of the electron gas in the surface region. This potential is of the same general type as image potentials and, for an electron with the Fermi energy, it is asymptotic to the classical image potential at a distance x_0 of the order of 2 Å for sodium. This is in good agreement with the value obtained from the other calculations mentioned above. To calculate the work function it is necessary to consider electrons having the Fermi energy, although such electrons cannot be emitted from the metal. The exchange-correlation potential depends on the electron momentum, so that the surface barrier for an electron having a finite probability of escaping from the metal is slightly different. By extrapolating Bardeen's results it may be shown that $x_0 \sim 3{\cdot}3$ Å for an electron passing over the barrier.

---

† Now at The Research Laboratories, The General Electric Company Limited, Wembley, Middlesex.

The work function of an uncharged metal may be identified with the total work required to remove an electron from the metal to a large distance. The portion of this work contributed by classical image forces for a smooth plane surface is

$$\begin{aligned} \boldsymbol{e} V_1 &= \int_{x_0}^{\infty} F_1\,dx = \boldsymbol{e}/4x_0 \\ &= 1{\cdot}8\,\text{ev} \quad \text{for} \quad x_0 = 2\,\text{Å}. \qquad \ldots\ldots(2) \end{aligned}$$

The image force depends on the distribution of the induced surface charge, so that for a plane surface having microscopic irregularities, equation (1) does not hold for all distances x. At large distances from the surface the image force is not affected by the microgeometry as the charge is induced over an area large compared with the individual irregularities. Similarly, at sufficiently small distances equation (1) should be a good approximation since the atomic structure limits the smallest surface irregularity. It therefore seems unlikely that the image force can be modified within 2 Å of a surface : the failure of the classical law at x_0 also confirms this. At intermediate distances from an irregular surface the classical image force differs from that of a smooth surface. An electron outside a surface projection experiences a reduced image force because of the increased mean square distance to the surface charge; opposite a surface hollow the image force is increased. In these cases there is respectively a decrease and an increase in the image potential (equation (2)) for an electron moving from x_0 and leaving the surface.

The reduction of the image potential energy ($\boldsymbol{e}^2/4x_0$ for an electron) due to a surface projection has been calculated by Lewis (1954) for an idealized model consisting of a single prolate spheroidal boss on an infinite plane. Lewis shows that a reduction of up to 34% (0·61 ev for $x_0 = 2$ Å) is possible for the sharpest projection, roughly of this shape, likely to be found in practice. It is shown below that even with this reduction of the image potential, the work function of the metal at a surface projection is unaffected.

§ 2. THE WORK FUNCTION AND PATCH FIELDS

It may be shown that the work function of a metal has both volume and surface contributions. The volume components are determined entirely by the potentials of electrons in the interior of the metal, and include the total work done against all the image-type forces : the surface shape can only affect the rate of the change of this potential on moving away from the metal. The true surface component of the work function is the change of potential $\boldsymbol{e}D$ due to a surface double layer of D volts. In a real metal the largest contribution to the surface double layer is due to orientation of the molecules of an adsorbed gas or to a surface layer of ions, but there is also an inherent double layer on an ideal metal surface due to the extension of the electron gas beyond the surface cells of the metal. Bardeen (1936 b) showed that this accounted for about one sixth of the work function of sodium. The second possible component to the natural double layer of an ideal surface is due to the lateral spreading of the electron cloud from projections of the surface cells of the metal (Smoluchowski 1941). Both of these components may be assumed to be independent of the surface microgeometry including irregularities made up of several atomic cells. In particular, the calculations showed that the spread of electrons is insensitive to the shape of the potential barrier even at small distances. The change of image potential outside a surface irregularity only occurs at distances of several times x_0. From this it can be seen that the surface barrier of a uniform metal comprises a dipole layer and other short range forces within about x_0 and the image force alone at greater distances.

The potential barrier of a non-uniform surface includes another long range electrostatic force. The non-uniformities usually considered are due to large patches of adsorbed gas changing the dipole moment, and so the work function, of regions of the surface. The total work required to remove an electron from the metal to a very large distance must obviously be independent of its path through the surface, and to compensate for variations of work function on the surface, long range 'patch fields' are set up. These may be thought of as the contact potential fields linking regions of high and low work function on a single surface. Like contact potentials, they are associated with a surface charge distribution. The complete surface potential is shown in figure 1. The potential of the patch field depends entirely on the distribution and size of the surface patches: usually it is only appreciable far outside the range of image forces (at 100 Å the image potential is 0·036 ev and negligible for many purposes). It is then possible to define approximately a 'local work function' ϕ_i as the work required to move an electron far enough from the metal for the image potential to be negligible. Then it may be shown that the work required to move an electron to infinity is

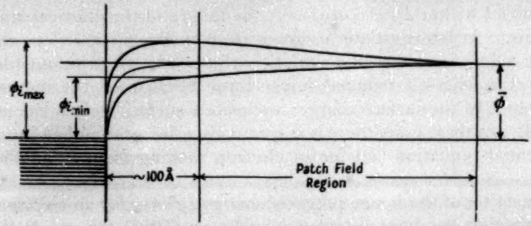


Figure 1. Typical surface potential barrier for a metal having large non-uniformities.

$$\bar{\phi}=\sum_i f_i\phi_i \qquad \ldots\ldots(3)$$

where f_i is the fraction of the surface covered by patches of type i (Herring and Nichols 1949).

The non-uniformity to be considered in the present case is the surface geometry. As shown above, it is reasonable to assume that for a clean surface the microgeometry affects only the image potential. As shown by Lewis (1954) this is less for an electron moving from x_0 to about 100 Å outside a surface projection than it is for a plane surface. The lowering of the surface barrier is compensated, for an electron moving to larger distances, by the patch field of the projection. If this patch field extended much more than about 100 Å from the surface it would be possible, as in figure 1, to speak of a local work function which would be less than the work function of a smooth surface. It is shown below, however, that in fact the patch field is so close to the surface that it compensates for the lowering of the image potential within the range of image forces.

§ 3. THE PATCH FIELD OF AN IRREGULAR SURFACE

The method of calculating the patch potential for an irregular surface follows the same principles as for a surface having large-area changes of double layer (e.g. Becker 1935). There are, however, the following complications : (*a*) the patch field is much more concentrated and largely within the image region, (*b*) the factor modifying the work function (the change of image potential) occurs only at distances larger than x_0 from the surface, (*c*) the geometry is necessarily complex. The stages in the calculation are as follows:

(i) The image potential (corresponding to equation (2) for a smooth surface) for an electron moving from x_0 to infinity, has to be found for each point on the surface. Near convex parts of the surface the value would be less than that of equation (2) and at concave parts, greater. The mean value would be $e/4x_0$, indicating that the average work function is the same as that of a smooth surface. This follows from equation (3) and the fact that the true surface component eD cannot be affected by microgeometry.

(ii) The electrostatic potential of each surface element follows from the difference of the image potential calculated at (i) and the value $e/4x_0$ for a smooth surface.

(iii) The shape of the patch field can, in principle, be found by setting up the surface potentials of (ii).

(iv) The comparison of the change of image potential with the patch potential relative to the surface gives the net change of potential at all points in the space outside the surface so that the shape of the surface potential barrier can be found for all paths through the surface.

The distribution of electrostatic potential (and so of charge) over any irregular or non-uniform surface is necessary for the Fermi level to be the same everywhere in the metal. It may be shown that this corresponds to the correct equilibrium condition, i.e. that the electrochemical potential for electrons is constant. This is more general than the classical electrostatic requirement of a conducting surface being an equipotential. It is for this reason that electrostatics does not account for contact potentials. The charge distribution giving rise to the patch field can only be in the surface, although the change of image potential occurs at larger distances than x_0. This is the only way in which the potential energies of electrons can be constant at all points in the metal and at all distant points.

The patch potential arising from a change of image potential must be regarded as a volume component to the work function of an irregular metal since the total work done against patch plus image potentials is the same as for a smooth surface.

§ 4. DETERMINATION OF PATCH POTENTIAL

The procedure outlined above is far too difficult to carry out even for the idealized model of a surface irregularity as used by Lewis. An analytical solution for the image potential can be found only along the axis of the spheroidal boss representing a surface projection. Even if the correct surface potential could be set up, the calculation of the patch field would not be easy with this geometry. Accordingly, an approximate solution has been obtained for the model using an electrolytic tank.

Figure 2 shows the approximate shape of the patch field due to a surface charge distribution based on the probable change of image potential for each point. The electrostatic potential at the vertex of the boss is $+V_P$ corresponding to the lowering of the image potential calculated by Lewis. The largest increase of image potential occurs where the boss meets the plane, and it gives a potential $-V_Q$ there. As the available calculation is for the image potential along the axis of the boss, it is only required to find the patch potential along the same line. It is found that this varies approximately as $1/x$ and that it depends very little on the potential distribution on the sides of the boss.

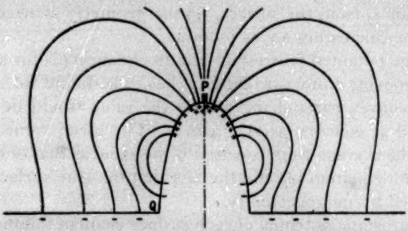


Figure 2. Approximate shape of patch field for the model of a surface projection.

The experimental determination of the three-dimensional patch potential was carried out in the electrolytic tank for a model of a typical spheroid used by Lewis. Due to lack of information about the image potential away from the axis, the potential distribution had to be estimated as in figure 2 except for the vertex of the boss. A further approximation on the model was the replacement of the continuous potential distribution by a stepped distribution fixed by a series of equipotential foils as shown in figure 3. This distorts the field only near the gaps between foils and scarcely at all along the axis. The axial field shown in figure 4 was obtained for foil potentials (figure 3) giving a more gradual potential distribution on the sides of the boss than would be expected, so that it probably underestimates the axial field. The potential at the base of the spheroid should ideally be negative to allow for the increased image potential, but this requirement also does not affect the axial field.

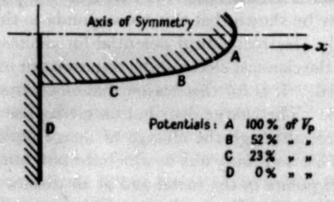


Figure 3. Surface potential distribution for electrolytic tank model of projection.

The modification to the image potential, giving a lowering eV_P of the work function, for the axis of a boss of the same size was obtained by interpolation from Lewis' paper and compared with the compensating effect of the patch potential. The result, figure 4, shows that the two effects exactly cancel for all but very small distances from the surface. Thus the work required to extract an electron 50 Å from a surface projection is the same as that to extract it from a plane surface. Within this distance the accuracy of the solution falls, but it is possible that there is a small resultant hump added to the surface barrier of the projection.

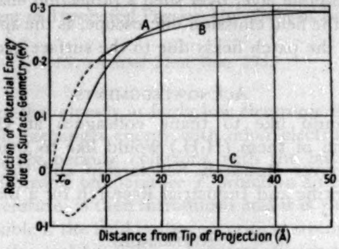


Figure 4. Components of change of potential energy along the axis of a surface projection 60 Å high and of 30 Å base diameter: A, reduction of image potential; B, compensating patch potential; C, net change of potential energy.

§ 5. CONCLUSIONS

It has been shown that the intensity of the patch field of an irregular metal surface has hitherto been underestimated (Lewis 1954). The lowering of the image potential for an electron outside a surface projection is compensated well within the image region, by a patch field, so that the shape of the surface barrier is unaffected. This result has been proved explicitly only for a particular model of a surface projection. The specific calculation is for a surface irregularity giving a moderate lowering (0·29 ev) of the image potential, but since the result appears to be conclusive it has been thought unnecessary to repeat it for projections at different sizes. It has been shown that the major assumptions involved in the calculation of a patch field within the range of image forces are connected with the choice of the charge distribution on the sides of the surface projection. However, these scarcely affect the result for the potential outside the vertex of the boss for which detailed calculations of image potential are available.

Electron emission from a non-uniform metal surface in accelerating fields is well known to show large deviations from the ideal Schottky form because of the change of surface barrier with patch fields (Herring and Nichols 1949). When the differences of work function over the surface are due to surface microgeometry the patch field is concentrated so close to the surface, and it so exactly balances the lowering of the image potential, that the shape of the surface barrier is the same as for an ideal surface, and electron emission should obey the Schottky equation. Indeed, if this were not so, it is unlikely that Schottky plots agreeing with the theory for ideal surfaces would have been obtained except with extreme precautions, because surface projections even 10 Å high have a reduced image potential, and changes of work function over a very small area can be detected by electron emission. There is thus experimental evidence that the patch field for surface irregularities is concentrated within the image region.

The main effect of surface geometry on electron emission is the enhancemen of an applied electric field. This is unlikely to affect the characteristics o Schottky emission in fields of less than 10^4 v cm^{-1} (Morant 1955), although fo higher fields (particularly in the field emission range) it becomes increasingly important, so that most of the emission then comes from surface projections Changes of surface double layer over such a projection enable the surface struc ture to be studied in the field emission microscope, as the applied fields are locally much stronger than the patch fields due to the surface geometry and variations of double layer.

ACKNOWLEDGMENTS

The authors would like to thank colleagues at Queen Mary College for discussions. One of them (H.H.) would like to thank the University o London for the Hoover Research Scholarship and the other (M.J.M.) the Department of Scientific and Industrial Research for a maintenance allowance

REFERENCES

BARDEEN, J., 1936 a, *Phys. Rev.*, **49,** 640; 1936 b, *Ibid.*, **49,** 653; 1940, *Ibid.*, **58,** 727.
BECKER, J. A., 1935, *Rev. Mod. Phys.*, **7,** 95.
HERRING, C., and NICHOLS, M. H., 1949, *Rev. Mod. Phys.*, **21,** 185.
JURETSCHKE, H. J., 1953, *Phys. Rev.*, **92,** 1140.
LEWIS, T. J., 1954, *Proc. Phys. Soc.* B, **67,** 187.
MORANT, M. J., 1955, *Proc. Phys. Soc.* B, **68,** 513.
SACHS, R. G., and DEXTER, D. L., 1950, *J. Appl. Phys.*, **21,** 1304.
SMOLUCHOWSKI, R., 1941, *Phys. Rev.*, **60,** 661.

Dielectric Breakdown in Solids †

By H. FRÖHLICH and B. V. PARANJAPE
Department of Theoretical Physics, University of Liverpool

MS. received 22nd June 1955

Abstract. It is shown that already at fairly low electronic densities (of the order 10^{14} cm^{-3}) an electron exchanges energy with other electrons quicker than with lattice vibrations, mainly because collisions with the latter are nearly elastic. The notation of an electronic temperature T, which in an external field is higher than the lattice temperature, is then introduced and it is shown that a stationary state becomes impossible if the field is above a critical strength F_c the breakdown strength. For nonpolar and for polar substances F_c can be expressed in terms of the electronic mobility in weak fields, and of other measurable quantities.

§ 1. INTRODUCTION

1.1. *General*

The application of moderately strong electric fields to solid insulators leads usually to non-ohmic currents. Rose (1955) has shown recently that often these currents can be interpreted as space charge limited currents; the electric field between the electrodes is then not homogeneous even under excellent geometrical conditions. With increasing field strength occurrence of internal ionization should lead to the formation of temporary electron avalanches whose aspects have been discussed by Seitz (1949; for their relation to breakdown see however Fröhlich and Seitz 1950). In certain circumstances this process may, at sufficiently strong fields, lead to a breakdown of the solid. It must be emphasized that in this type of breakdown the critical field strength must depend on the thickness of the sample. It differs, therefore, from the so-called intrinsic breakdown in which the critical field strength F_c is independent of thickness (e.g. Whitehead 1951). In this case of intrinsic breakdown, therefore, the electric field F should be homogeneous (assuming appropriate geometry) even if F is slightly below the breakdown strength F_c. One may expect that in this case the actual occurrence of breakdown is subject to much smaller fluctuations than avalanche breakdown and it is of interest, therefore, that Kawamura and his collaborators (1954) seem to have been able to observe experimentally these two types of breakdown.

The present paper is solely concerned with intrinsic breakdown. Although here, too, the actual breakdown of the lattice occurs through a narrow channel, the independence of F_c of thickness must mean that in this case the condition for the instability is not connected with the formation of avalanches. Let $j(F)$ be the electric current density at the field strength F, and assume that the heat conductivity of the lattice is sufficiently great to conduct away from each volume element $d\tau$ the energy $B\,d\tau$ transferred from the electrons to the lattice[†]. Hence the condition

---

† Partly based on reports L/T 277, 285, 307 of the British Electrical and Allied Industries Research Association.

$$A(F)>B, \quad A(F)=j(F)F, \quad \text{if} \quad F>F_c \qquad \ldots\ldots(1.1)$$

implies that in each volume element $d\tau$ the electrons receive energy from the field at a higher rate than they can transfer it to the lattice. When this condition is fulfilled a stationary state becomes impossible. The condition does not, however, indicate the way in which the unstable situation develops until breakdown occurs. Condition (1) was first introduced by Fröhlich (1937) for a special model; it is essentially different from breakdown conditions in gases or from non-intrinsic breakdown in solids which usually deal with non-homogeneous situations.

For an appropriate theoretical treatment of dielectric breakdown it would be desirable to calculate the electronic momentum distribution in the presence of a homogeneous external field, and to derive the conditions on which a stationary state is impossible. It has been shown previously (Fröhlich 1947 b) that the simple model of free electrons interacting with the lattice vibrations only can never lead to a stationary distribution, mainly because the mean free path of the electrons increases with energy. If other processes, like inelastic collisions with bound electrons or collisions with other free electrons, are included then stationary conditions can be obtained in sufficiently weak fields. Several attempts have been made to introduce internal ionization and recombination processes but they are all open to objections (in some of the papers ionization processes are introduced without consideration of the inverse processes!).

A possibility for a fairly simple way of finding a solution arises when the density of electrons is so high that the rate of energy exchange between electrons through mutual collisions is large compared with the rate of energy exchange between electrons and lattice vibrations. In this case it may be assumed that the electrons are in thermal equilibrium (though displaced in momentum space) at a temperature T which may be higher than the lattice temperature T_0. This assumption was first introduced by Fröhlich (1947 a) for the case of imperfect crystals containing many shallow electron traps. In the present paper it is intended to investigate the case of pure crystals containing a negligible number of traps.

1.2. *Range of Validity of the Model*

At first sight it might seem that a very high density of free electrons (i.e. electrons in conduction levels) is required to make interelectronic collisions more effective than collisions between electrons and lattice vibrations. That this is not true is mainly due to the small energy exchange for the latter collisions. Thus in a non-polar lattice with sound velocity s, momentum conservation requires that the average energy exchange per collision (i.e. the energy of a lattice quantum), is of the order $ps\simeq(Ems^2)^{1/2}$ where p is the momentum and E the kinetic energy of the electron. Also the ratio of the probabilities of emission and of absorption of a lattice quantum by an electron is $(1+\bar{n})/\bar{n}$ where $\bar{n}\simeq kT_0/(Ems^2)^{1/2}$ is the average number of excited quanta of a normal mode which can interact with an electron of energy E (provided $\bar{n}>1$ which holds in a wide range). If now $\tau(E, T_0)$ is the average time between two collisions, only one in $1+2\bar{n}$ collisions leads to a net energy loss so that

---

[†] If this does not hold then yet another type of breakdown (through thermal instability) may occur (cf. Whitehead 1951).

$$\left(\frac{dE}{dt}\right)_{\mathrm{L}} \simeq -\frac{(Ems^2)^{1/2}}{\tau(E, T_0)}\,\frac{(Ems^2)^{1/2}}{kT_0} = -\frac{E}{kT_0}\,\frac{ms^2}{\tau(E, T_0)} \qquad \ldots\ldots(1.2)$$

represents the order of magnitude of the rate of energy transfer from electrons of energy E to the lattice, provided that the average E is larger than kT_0. Thus electrons which have gained energy beyond their equilibrium value lose it to the lattice at the very low rate of about ms^2 per collision (provided E is of the same order as kT_0), as is quite well known (note that $ms^2/k \simeq 1\,°\mathrm{K}$).

The rate of loss of energy of such an electron due to collisions with other electrons is of the order (e.g. Pines 1953)

$$\left(\frac{dE}{dt}\right)_{\mathrm{e}} = -\frac{4\pi n e^{*4}}{p}\log\gamma \qquad \ldots\ldots(1.3)$$

where e^* is an 'effective' electronic charge differing from e through effects of the dielectric constant of the material (e^{*2} usually decreases with increasing velocity). The logarithmic factor depends on the density n of electrons and on their velocity; in the range of values required here it is of order unity. From (1.2) and (1.3) it follows that the density n_0 at which the two rates of energy transfer are equal is of the order

$$n_0 \simeq \frac{1}{4\pi}\,\frac{E^{3/2}m^{3/2}s^2}{kT_0 e^{*4}\tau}. \qquad \ldots\ldots(1.4)$$

For n greater than n_0 the assumption of an electronic temperature T should be valid. If E is larger than, but of the same order as, kT_0 then at room temperature, with $\tau \sim 10^{-14}$ second, $e^{*4}/e^4 \simeq 10$ and m equal to the electronic mass, n is of the order $10^{14}\ \mathrm{cm}^{-3}$. Though n_0 increases with the average electronic energy E which of course increases with field strength, the assumption of an electronic equilibrium at a temperature T greater than T_0 should reasonably well be valid in many cases, in particular if the electronic density at breakdown is of the order $10^{17}\ \mathrm{cm}^{-3}$ as estimated by O'Dwyer (1954).

In typical ionic crystals it is usually assumed that the electrons interact with the optically active vibrations only which have all approximately the same frequency ω (e.g. Fröhlich 1954). If $kT \ll \hbar\omega$ most electrons can not emit quanta. Scattering is then best described as elastic in terms of a two-stage process, absorption and re-emission of quanta $\hbar\omega$. Clearly no energy transfer at all takes place in this approximation. The only remaining energy transfer is then due either to electrons with energy E greater than $\hbar\omega$ or due to interaction with acoustic waves which is assumed to be very weak. Hence if the treatment of ionic crystals described above is a good approximation, then interelectronic collisions should already be of importance at very low electronic densities. Their value would probably depend on the strength of the interaction between electrons and the acoustic vibrations.

If the condition discussed above for rapid energy exchange amongst electrons is fulfilled then it can be assumed that interelectronic collisions essentially determine the electronic distribution function $f(\mathbf{p})$ which must thus be a Maxwell distribution which may be displaced in momentum space, i.e.

$$f(\mathbf{p}) = a\exp\left(-|\mathbf{p}-\mathbf{p}_0|^2/2mkT\right), \quad \int f(\mathbf{p})\,d^3p = N \qquad \ldots\ldots(1.5)$$

where a is independent of p but is determined by the total number N of electrons.

This function $f(\mathbf{p})$ contains two parameters p_0 and T which will be determined from the rate of momentum and energy exchange with the lattice. The momentum displacement $\mathbf{p}_0$ will be found proportional to the electric field, and the condition

$$\frac{p_0^2}{2mkT} \ll 1 \qquad \ldots\ldots(1.6)$$

will hold throughout this paper. Hence if $f(\mathbf{p})$ is developed in spherical harmonics with the field direction as axis,

$$f(\mathbf{p}) = f_0(p) + f_1(p)\mathrm{P}_1(\cos\theta) + \ldots \qquad \ldots\ldots(1.7)$$

where θ is the angle between p and the field, and approximately ($f_1(p) = -p_0\partial f_0/\partial p$ is an exact relation)

$$f_0(p) \simeq a\exp(-p^2/2mkT) \qquad \ldots\ldots(1.8)$$

$$f_1(p) = -p_0\frac{\partial f_0(p)}{\partial p} \simeq \frac{pp_0}{mkT}f_0(p). \qquad \ldots\ldots(1.9)$$

It must be mentioned that the condition $n > n_0$ for strong interelectronic energy exchange justifies the form (1.8) of $f_0(p)$ only. To obtain (1.9) for $f_1(p)$ requires that the rate of momentum exchange amongst electrons be large compared with that between electrons and lattice vibrations, a condition which requires much higher electron densities than n_0. Nevertheless (1.9) is always a reasonably good approximation for $f_1(p)$, because the usual method in which interelectronic collisions are neglected makes f_1 proportional to $\tau(p)\partial f_0/\partial p$. The electric current calculated in this way then differs from the one obtained from (1.9) only by the way of averaging the relaxation time $\tau(p)$ over p. This leads to a slightly different numerical factor which is not significant for the purpose of the present paper.

Clearly the method discussed here can be generalized to include other modes of loss of energy from the electron system than collisions with the lattice vibrations. Amongst these emission of quanta by capture of electrons into shallow traps has been treated previously but is excluded from the present paper, perfect crystals being assumed. A possible recombination with vacancies connected with simultaneous emission of many quanta cannot be treated at present because very little is known about this process. Internal ionization processes by fast electrons, no doubt, are connected with absorption or emission of single quanta. Both these processes are unlikely, however, to be of importance for the total energy balance. The present treatment, therefore, should hold in a fairly general manner if $n > n_0$. This condition is almost certainly fulfilled for breakdown in semiconductors (e.g. Ge, Si). It is difficult to say in which range (of temperature, crystallinity, etc.) it holds in the case of insulators but, as mentioned above, this might be decided experimentally by measurements of fluctuations. Whenever the condition $n > n_0$ is not fulfilled then a theory in which interelectronic collisions are neglected (e.g. Fröhlich 1937) should hold. This leads, in general, to a lower breakdown strength than is found in the present model. The latter, therefore, should always give an upper theoretical limit.

§ 2. CALCULATIONS

Consider the electrons in the conduction band as free (with effective mass m) except for mutual collisions and for collisions with the lattice vibrations. Then the distribution function $f(p)$ in momentum space satisfies the kinetic equation

$$\frac{\partial f}{\partial t} = \left(\frac{\partial f}{\partial t}\right)_{\mathrm{F}} + \left(\frac{\partial f}{\partial t}\right)_{\mathrm{L}} + \left(\frac{\partial f}{\partial t}\right)_{e} \qquad \ldots\ldots(2.1)$$

where the terms with suffixes F, L and e respectively denote the rates of change of f due to an external field, to collisions with the lattice vibrations, and to the interelectronic collisions. In the stationary state

$$\frac{\partial f}{\partial t}=0. \qquad \ldots\ldots(2.2)$$

Since interelectronic collisions conserve energy and momentum it follows that

$$\sum_{\mathbf{p}} \mathbf{p}\left(\frac{\partial f}{\partial t}\right)_{\mathrm{e}}=0 \quad \text{and} \quad \sum_{\mathbf{p}} E_p\left(\frac{\partial f}{\partial t}\right)_{\mathrm{e}}=0 \qquad \ldots\ldots(2.3)$$

where

$$E_p=p^2/2m. \qquad \ldots\ldots(2.4)$$

Hence from (2.1)–(2.3),

$$\sum_{\mathbf{p}} \mathbf{p}\left\{\left(\frac{\partial f}{\partial t}\right)_{\mathrm{F}}+\left(\frac{\partial f}{\partial t}\right)_{\mathrm{L}}\right\}=0 \qquad \ldots\ldots(2.5)$$

and

$$\sum_{\mathbf{p}} E_p\left\{\left(\frac{\partial f}{\partial t}\right)_{\mathrm{F}}+\left(\frac{\partial f}{\partial t}\right)_{\mathrm{L}}\right\}=0. \qquad \ldots\ldots(2.6)$$

Moreover as discussed in § 1 the distribution function is of the form (1.5) and thus contains two parameters p_0 and T which can be obtained from the two equations (2.5) and (2.6).

Let the field F be in the z direction so that

$$\frac{dp_z}{dt}=eF, \quad \frac{dp_x}{dt}=\frac{dp_y}{dt}=0. \qquad \ldots\ldots(2.7)$$

Then

$$-\left(\frac{\partial f}{\partial t}\right)_{\mathrm{F}}=\frac{\partial f}{\partial p_z}\frac{dp_z}{dt}=eF\frac{\partial f}{\partial p_z} \qquad \ldots\ldots(2.8)$$

and hence with

$$\sum_{\mathbf{p}} \rightarrow \int d^3p,$$

$$\int p_z\left(\frac{\partial f}{\partial t}\right)_{\mathrm{F}} d^3p=-eF\int p_z\frac{\partial f}{\partial p_z}\,d^3p=eFN \qquad \ldots\ldots(2.9)$$

making use of (1.5). Also

$$\int E_p\left(\frac{\partial f}{\partial t}\right)_{\mathrm{F}} d^3p=-\frac{eF}{2m}\int p^2\frac{\partial f}{\partial p_z}\,d^3p=\frac{eF}{m}\int p_z f\,d^3p=j(F)FN \qquad \ldots\ldots(2.10)$$

which with (1.7)–(1.9) becomes ($p_z=p\cos\theta$)

$$\int E_p\left(\frac{\partial f}{\partial t}\right)_{\mathrm{F}} d^3p=-\frac{eF}{m}p_0\int p\frac{\partial f_0}{\partial p}\cos^2\theta\,d^3p=\frac{eFp_0N}{m}. \qquad \ldots\ldots(2.11)$$

In order to calculate $(\partial f/\partial t)_{\mathrm{L}}$, the processes of absorption and emission of quanta of lattice vibrations by the electrons must be considered. Let $\mathbf{q}/\hbar$ be the wave vector of a quantum, and let $\hbar\omega(q)$ be its energy. Then conservation of momentum requires that electronic transitions from a state with momentum $\mathbf{p}$ are possible only into the state $\mathbf{p}+\mathbf{q}$ in the case of absorption, and $\mathbf{p}-\mathbf{q}$ in the case of emission of a quantum. Let $\rho_{\mathrm{a}}(\mathbf{p},\mathbf{p}+\mathbf{q})$ and $\rho_{\mathrm{e}}(\mathbf{p},\mathbf{p}-\mathbf{q})$ respectively be the probabilities per second for these transitions. Then clearly

$$\left(\frac{\partial f(\mathbf{p})}{\partial t}\right)_{\mathrm{L}}=-\sum_{\mathbf{q}}\{f(\mathbf{p})\rho_{\mathrm{a}}(\mathbf{p},\mathbf{p}+\mathbf{q})-f(\mathbf{p}+\mathbf{q})\rho_{\mathrm{e}}(\mathbf{p}+\mathbf{q},\mathbf{p})+f(\mathbf{p})\rho_{\mathrm{e}}(\mathbf{p},\mathbf{p}-\mathbf{q})$$
$$-f(\mathbf{p}-\mathbf{q})\rho_{\mathrm{a}}(\mathbf{p}-\mathbf{q},\mathbf{p}). \qquad \ldots\ldots(2.12)$$

The transition probabilities ρ can be calculated by standard quantum mechanical procedure leading to

$$\rho_{\mathrm{a}}=\frac{2\pi}{\hbar}B(q)n_q\delta(y_{\mathrm{a}}) \qquad \ldots\ldots(2.13)$$

and

$$\rho_{\mathrm{e}}=\frac{2\pi}{\hbar}B(q)(1+n_q)\delta(y_{\mathrm{e}}) \qquad \ldots\ldots(2.14)$$

with

$$y_{\mathrm{a}}=E_{\mathrm{f}}-E_{\mathrm{i}}-\hbar\omega(q); \quad y_{\mathrm{e}}=E_{\mathrm{f}}-E_{\mathrm{i}}+\hbar\omega(q). \qquad \ldots\ldots(2.15)$$

Here $B(q)$ is the square of the interaction matrix element, E_i and E_f respectively are the energies of the initial and final electronic states and

$$n_q = [\exp(\hbar\omega(q)/kT_0) - 1]^{-1}, \quad \text{i.e. } 1 + n_q = n_q \exp(\hbar\omega(q)/kT_0) \qquad \ldots\ldots(2.16)$$

Now in a system of polar coordinates with $\mathbf{p}$ as axis

$$\sum_{\mathbf{q}} = \frac{V}{(2\pi)^3\hbar^3}\int q^2 dq\, d\cos\beta\, d\psi \qquad \ldots\ldots(2.17)$$

where ψ is the azimuth, β is the polar angle and V is the total volume in configuration space. Integration over β can then be used to eliminate the δ-functions This enforces energy conservation, i.e.

$$\frac{|\mathbf{p}\pm\mathbf{q}|^2}{2m} = \frac{p^2}{2m} + \hbar\omega(q) \qquad \ldots\ldots(2.18)$$

so that the angles between $\mathbf{p}$ and $\mathbf{q}$ are given by

$$\cos(\mathbf{pq})_{\pm} = \mp\frac{q}{2p} + \frac{m\hbar\omega(q)}{pq}. \qquad \ldots\ldots(2.19)$$

Thus after integration over β, and making use of (2.13)–(2.19), equation (2.12) becomes

$$\left(\frac{\partial f(\mathbf{p})}{\partial t}\right)_{\mathrm{L}} = -\frac{V}{(2\pi)^2\hbar^4}\frac{m}{p}\int\frac{d\psi}{q}q^2 dq B(q) n_q$$
$$\times\{[f(\mathbf{p}) - f(\mathbf{p}+\mathbf{q})\exp(\hbar\omega(q)/kT_0)]$$
$$+[f(\mathbf{p})\exp(\hbar\omega(q)/kT_0) - f(\mathbf{p}-\mathbf{q})]\}. \qquad \ldots\ldots(2.20)$$

The vectors $\mathbf{q}$ appearing in (2.20) have polar angles given by (2.19).

Now making use of (1.5) with (2.18) and (2.19),

$$f(\mathbf{p}\pm\mathbf{q}) = f(\mathbf{p})\exp(\mp\hbar\omega(q)/kT_0)\exp(\pm\mathbf{p}_0\mathbf{q}/kT) \qquad \ldots\ldots(2.21)$$

or (cf. (1.6)) developing up to terms of first order in p_0, introducing (1.7)–(1.9),

$$f(\mathbf{p}\pm\mathbf{q}) \simeq f_0(p)\left(1 + \frac{p_0}{mkT}\{p\cos\theta \pm q\cos(\mathbf{p}_0\mathbf{q})\}_{\pm}\right)\exp(\mp\hbar\omega(q)/kT). \qquad \ldots\ldots(2.22)$$

Clearly (2.20) vanishes if $T = T_0$ and $p_0 = 0$, i.e. in the absence of an external field. It will be noticed that $\cos(\mathbf{p}_0\mathbf{q})$ is the only quantity depending on ψ, and that

$$\int_0^{2\pi}\cos(\mathbf{p}_0\mathbf{q})_{\pm}d\psi = \int\{\cos\theta\cos(\mathbf{pq})_{\pm} + \sin\theta\sin(\mathbf{pq})_{\pm}\cos\psi\}d\psi$$
$$= 2\pi\left(\mp\frac{q}{2p} + \frac{m\hbar\omega(q)}{pq}\right)\cos\theta. \qquad \ldots\ldots(2.23)$$

Thus from (2.20), using (2.22), (2.23) and (1.7)–(1.9),

$$-\left(\frac{\partial f}{\partial t}\right)_{\mathrm{L}} = g_0(p) + p_0\cos\theta\, g_1(p), \qquad \ldots\ldots(2.24)$$

where
$$g_0(p) = \frac{V}{2\pi\hbar^4}\frac{m}{p}f_0(p)\int q\,dq\,B(q)\,n_q$$
$$\times\left\{\left[1 - \exp\left(\frac{\hbar\omega(q)}{kT_0} - \frac{\hbar\omega(q)}{kT}\right)\right]_+ + \left[\exp\left(\frac{\hbar\omega(q)}{kT_0}\right) - \exp\left(\frac{\hbar\omega(q)}{kT}\right)\right]_-\right\} \qquad \ldots\ldots(2.25)$$

and

$$g_1(p)=\frac{V}{2\pi\hbar^4}\frac{f_0(p)}{kT}\int q\,dq\,B(q)\,n_q$$
$$\times\left\{\left[1-\exp\left(\frac{\hbar\omega(q)}{kT_0}-\frac{\hbar\omega(q)}{kT}\right)\right]_+ +\left[\exp\left(\frac{\hbar\omega(q)}{kT_0}\right)-\exp\left(\frac{\hbar\omega(q)}{kT}\right)\right]_-\right.$$
$$+\left(\frac{q^2}{2p^2}-\frac{m\hbar\omega(q)}{p^2}\right)_+\exp\left(\frac{\hbar\omega(q)}{kT_0}-\frac{\hbar\omega(q)}{kT}\right)$$
$$\left.+\left(\frac{q^2}{2p^2}+\frac{m\hbar\omega(q)}{p^2}\right)_-\exp\left(\frac{\hbar\omega(q)}{kT}\right)\right\}. \qquad \ldots\ldots(2.26)$$

The integration limits are partly determined by the fact that $|\cos(\mathbf{qp})_\pm|$ as given by (2.19) must be smaller than unity. This involves, in general, different upper and lower limits $q_{\pm u}$ and $q_{\pm l}$ for the two cases denoted by + or −.

It now follows from (2.14) that

$$\int p_z\left(\frac{\partial f}{\partial t}\right)_L d^3p=\frac{p_0}{3}\int p\,g_1(p)\,d^3p \qquad \ldots\ldots(2.27)$$

and

$$\int E_p\left(\frac{\partial f}{\partial t}\right)_L d^3p=\int\frac{p^2}{2m}g_0(p)\,d^3p. \qquad \ldots\ldots(2.28)$$

Hence, introducing (2.27), (2.28), (2.9) and (2.11) in to (2.5) and (2.4) leads to

$$eFN=\frac{p_0}{3}\int p\,g_1(p)\,d^3p \qquad \ldots\ldots(2.29)$$

and

$$eFNp_0=\tfrac{1}{2}\int p^2 g_0(p)\,d^3p. \qquad \ldots\ldots(2.30)$$

The first of these equations shows that p_0 is proportional to F, as is to be expected. The second equation can then be used to calculate the electronic temperature T in terms of F or p_0. It will be noticed that g_0 vanishes if $T=T_0$, but g_1 does not vanish in this limit. It must be pointed out here that (2.29) is of the first order in p_0 but, using this result, the left-hand side of (2.30) is of the second order. Now in the derivation of $(\partial f/\partial t)_L$ terms which are of second order in p_0 have been neglected. These terms would make an additional contribution to the right-hand side of (2.30). While the term with g_0 represents the rate of transfer of electronic energy to the lattice due to the temperature difference $T-T_0$, the missing terms represent the transfer due to the displacement of electronic distribution in momentum space. To neglect these terms would be wrong for some models in the case of intermediate fields where $T-T_0$ is not large, but is significant. For fields near breakdown, however, $T-T_0$ is so large that it is correct to neglect $p_0{}^2$ terms on the right-hand side of (2.30). For details of the proof the reader is referred to two papers by Paranjape (1953, 1954).

In order to calculate the electronic temperature, p_0 can be eliminated from (2.29) and (2.30) leading to

$$e^2F^2N^2=\tfrac{1}{6}\int p^2 g_0(p)\,d^3p\int p\,g_1(p)\,d^3p. \qquad \ldots\ldots(2.31)$$

This equation expresses that the electronic temperature T must be such that the rate of transfer of energy from the field to the electrons must be equal to the rate of transfer from the electrons to the lattice vibrations. From (2.25) and (2.26) it can be seen that the right-hand side of (2.31) vanishes for $T=T_0$ (because then g_0 vanishes) and for $T\to\infty$ (because then g_0 and g_1 vanish). Thus this quantity has a maximum for a certain $T=T_c$ which defines the maximum value F_c of the field for which (2.31) has a solution. For larger fields, therefore, a stationary state becomes impossible.

§ 3. SPECIAL CASES

Further development of (2.29) and (2.30) requires a knowledge of the functions $B(q)$. It will be necessary to separate the cases of polar and of non-polar crystals.

Non-polar crystals. Here for simple lattices $\hbar\omega(q)=qs$, and as for metals (cf. Bloch 1928)

$$B(q)=\frac{2}{9}\,\frac{C^2qs}{Ms^2n_1V} \qquad \ldots\ldots(3.1)$$

where C is a constant energy of the order 1 ev, M is the ionic mass, s is the velocity of sound and n_1 the number of ions per unit volume. The largest value which q can take is

$$q_0=\hbar(6\pi^2n_1)^{1/3}. \qquad \ldots\ldots(3.2)$$

The integration limits for q are obtained from (2.19) as indicated after (2.26). Since in all practical cases $\hbar\omega(q)=qs\ll p^2/2m$,

$$q_{\pm 1}=0, \qquad q_{\pm u}=\min(2p, q_0) \qquad \ldots\ldots(3.3)$$

meaning that the upper limit is the smaller of the two quantities $2p$ or q_0.

High and low lattice temperatures T_0 will now be introduced and denoted by cases *a* and *b* respectively. These are temperatures T_0 satisfying (Θ is the Debye temperature)

$$kT_0>q_0s=k\Theta \quad (3.4\,a) \qquad \text{or} \qquad kT_0<q_0s=k\Theta. \quad (3.4\,b)$$

Hence

$$n_q\simeq kT_0/qs \quad (3.5\,a) \qquad \text{or} \qquad n_q\simeq\exp(-qs/kT_0). \quad (3.5\,b)$$

Near breakdown it will be found that the electronic temperature T is so high that kT is of the order 1 ev. It is possible, therefore, to neglect T_0/T against unity even in the high temperature case. Carrying through these approximations in a systematic way one finds from (2.25) and (2.26) with (3.1) and (3.3),

$$g_0(p)=\frac{\pi}{12}\,\frac{C^2m}{\hbar M}\,\frac{f_0(p)}{kT}\begin{cases} q_0/p & \text{if } 2p>q_0 \\ 16p^3/q_0^3 & \text{if } 2p<q_0 \end{cases} \qquad \ldots\ldots(3.4\,a,b)$$

$$g_1(p)=\frac{\pi}{12}\,\frac{C^2}{\hbar Ms^2}\,\frac{f_0(p)}{q_0}\,\frac{T_0}{T}\begin{cases} q_0^2/p^2 & \text{if } 2p>q_0 \\ 16p^2/q_0^2 & \text{if } 2p<q_0 \end{cases} \qquad \ldots\ldots(3.5\,a)$$

and

$$g_1(p)=\frac{\pi}{15}\,\frac{C^2}{\hbar Ms}\,\frac{f_0(p)}{kT}\begin{cases} q_0^2/p^2 & \text{if } 2p>q_0 \\ 32p^3/q_0^3 & \text{if } 2p<q_0 \end{cases} \qquad \ldots\ldots(3.5\,b)$$

These formulae show that the rate of energy and momentum transfer to the lattice has a maximum for electrons with momentum $q_0/2$, i.e. with an energy $q_0^2/8m$ which is of the order 1 ev.

The calculation of the integrals in (2.31) is now straightforward. It is found that the right-hand side of (2.31), i.e. also F^2, has a maximum for a temperature $T=T_c$ for which $2{\cdot}5kT_c\simeq q_0^2/2m$ in the high temperature case (*a*), and for which about $0{\cdot}9kT_c=q_0^2/8m$ in the low temperature case (*b*). These values of $T=T_c$ lead to the highest possible fields F_c for which a stationary state is possible. It is found that (using $k\Theta=q_0s$)

$$eF_c\simeq 0{\cdot}4\,\frac{4\pi}{3}\,\frac{C^2m}{\hbar Ms}\left(\frac{ms^2}{k\Theta}\right)^{1/2}\left(\frac{T_0}{\Theta}\right)^{1/2} \quad \text{if} \quad T_0\gg\Theta \qquad \ldots\ldots(3.6\,a)$$

and

$$eF_c\simeq 0{\cdot}2\,\frac{4\pi}{3}\,\frac{C^2m}{\hbar Ms}\left(\frac{ms^2}{k\Theta}\right)^{1/2} \quad \text{if} \quad T_0\ll\Theta. \qquad \ldots\ldots(3.6\,b)$$

The factors 0·4 and 0·2 respectively were obtained by numerical evaluation of the integrals.

Ionic crystals. As usual (cf. Fröhlich 1954) interaction with the optical branch of the lattice vibrations only will be considered. Then $\omega(q)=\omega$ is independent of q, and

$$B(q)=\frac{2\pi e^2\hbar^3\omega}{Vq^2}\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon}\right) \qquad \ldots\ldots(3.7)$$

where ϵ and ϵ_∞ are the static and the high frequency dielectric constants respectively. From (2.19) the integration limits for q in (2.25) and (2.26) are found to be

$$q_{\pm l}=p-(p^2\pm 2m\hbar\omega)^{1/2}$$
$$q_{\pm u}=p+(p^2\pm 2m\hbar\omega)^{1/2} \qquad \ldots\ldots(3.8)$$

provided the square root is real. If it is imaginary (the case with $p^2<2m\hbar\omega$) both limits must be equal, i.e. there is no contribution to the integrals. This is obvious because an electron with energy less than $\hbar\omega$ cannot emit a quantum. The upper limit must not exceed q_0; this is insignificant in the present case because it will be found that at breakdown kT is not much larger than $\hbar\omega$, i.e. it is much smaller than $q_0{}^2/8m$, in contrast to the case of non-polar substances.

The integrals (2.25) and (2.26) with the use of (3.7) and (3.8) now lead to

$$g_0(p)=eF_0\bar{n}\frac{f_0(p)}{p}\left\{\phi(p^2)\left[1-\exp\left(\frac{\hbar\omega}{kT_0}-\frac{\hbar\omega}{kT}\right)\right]\right.$$
$$\left.+\phi(p^2-2m\hbar\omega)\left[\exp\left(\frac{\hbar\omega}{kT_0}\right)-\exp\left(\frac{\hbar\omega}{kT}\right)\right]\right\}, \qquad \ldots\ldots(3.9)$$

and

$$g_1(p)=\frac{eF_0}{m}\bar{n}\frac{f_0(p)}{kT}\left\{\phi(p^2)\left[1-\left(1+\frac{m\hbar\omega}{p^2}\right)\exp\left(\frac{\hbar\omega}{kT_0}-\frac{\hbar\omega}{kT}\right)\right]\right.$$
$$+\phi(p^2-2m\hbar\omega)\left[\exp\left(\frac{\hbar\omega}{kT_0}\right)-\left(1-\frac{m\hbar\omega}{p^2}\right)\exp\left(\frac{\hbar\omega}{kT}\right)\right]$$
$$\left.+\frac{(p^2+2m\hbar\omega)^{1/2}}{p}\exp\left(\frac{\hbar\omega}{kT_0}-\frac{\hbar\omega}{kT}\right)+\frac{(p^2-2m\hbar\omega)^{1/2}}{p}\exp\left(\frac{\hbar\omega}{kT}\right)\right\}. \qquad \ldots\ldots(3.10)$$

Here

$$\phi(p^2)=\log\frac{(p^2+2m\hbar\omega)^{1/2}+p}{(p^2+2m\hbar\omega)^{1/2}-p} \qquad \ldots\ldots(3.11)$$

and $(p^2-2m\hbar\omega)^{1/2}$ has to be replaced by zero if $p^2<2m\hbar\omega$. F_0 is an electric field defined by

$$eF_0=\frac{me^2\omega}{\hbar}\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon}\right) \qquad \ldots\ldots(3.12)$$

and

$$\bar{n}=[\exp(\hbar\omega/kT_0)-1]^{-1}. \qquad \ldots\ldots(3.13)$$

Calculation of the integrals in (2.31) is straightforward now, though it requires some numerical work. The value kT_c of kT at which the right-hand side of (2.31) has its maximum, is found to be larger than $\hbar\omega$ though of the same order. Thus if $kT_0=\hbar\omega$, then $kT_c\simeq 3\hbar\omega$; if $T_0\rightarrow 0$ then $kT_c\simeq 2{\cdot}5\hbar\omega$.

The breakdown strength can be expressed as

$$F_c\simeq 0{\cdot}8F_0G(kT_0/\hbar\omega) \qquad \ldots\ldots(3.14)$$

where the function $G(x)$ ranges from $G(0)\simeq 2/3$ to $G(\infty)\simeq 1$.

Finally it should be mentioned that condition (1.6) is extremely well fulfilled in the case of non-polar crystals where at breakdown $p_0^2/2mkT_c$ is of the order 10^{-3}. For ionic crystals the value of this quantity is about 0·15 if $kT_0 = \hbar\omega$ and about 0·3 if $T_0 \to 0$.

§ 4. DISCUSSION

The results obtained in this paper depend on the possibility of rapid interchange of energy between conduction electrons through interelectronic collisions as discussed in § 1. It has been shown there that this condition is likely to be fulfilled for electronic densities met at fields near breakdown strength, mainly because energy transfer to the lattice vibrations is a slow process because their collisions with electrons are nearly elastic. The energy content of the electrons can then be described in terms of an electronic temperature T which increases with increasing field strength F. A critical field F_c exists, such that for F greater than F_c the electrons can no longer transfer energy to the lattice at the same rate at which they receive it from the field, i.e. equation (1.1) holds. It will be noticed using (1.7), (1.9), (2.11), (2.6), (2.29) and (2.30), that since $j(F) = ep_0N/mV$, equation (2.31) is identical with $A(F) = B$ as discussed in (1.1), if $j(F) = \sigma F$ is introduced, where $\sigma(T)$ is a conductivity which depends on the electronic temperature. The right-hand side of (2.31) is therefore essentially equal to B/σ. Its maximum value defines F_c by (2.31), and for $F > F_c$ therefore $A(F) > B$. The increase in electronic energy with time, in this case, must lead to a breakdown but the theory does not describe the manner in which it takes place. The electronic temperature at which this occurs is such that the magnitude of electronic energy kT is of the order $\hbar\omega \simeq 0{\cdot}1$ ev for ionic crystals and $q_0^2/8m \simeq 1$ ev for non-polar crystals.

The absolute magnitude of the breakdown† strength as is calculated in § 3 depends on the interaction between electrons and lattice vibrations only. It should be possible therefore, to express it in terms of the mobility which can be measured directly. Thus let $\mu(T_0)$ be the electronic drift mobility at temperature T_0 in weak fields. Then in non-polar crystals

$$\mu(T_0) = \frac{e}{m}\bar{\tau} = \frac{e}{m}\tau_0\left(\frac{\theta}{T_0}\right)^{3/2} \qquad \ldots\ldots(4.1)$$

where (e.g. Seitz 1948)

$$\frac{1}{\tau_0} = (2\pi)^{3/2}\frac{C^2m^{3/2}s}{\hbar M(k\Theta)^{3/2}}. \qquad \ldots\ldots(4.2)$$

Hence if a field F_1 is defined by (3.6 *a*, *b*) through

$$F_c = F_1(T_0/\Theta)^{1/2} \quad \text{if} \quad T_0 \gg \Theta; \qquad F_c = \tfrac{1}{2}F_1 \quad \text{if} \quad T_0 \ll \Theta \qquad \ldots\ldots(4.3)$$

then using (4.2), (3.2) and $k\Theta = q_0s$,

$$eF_1 = 0{\cdot}4\frac{4\pi}{3}\frac{C^2m^{3/2}}{\hbar M(k\Theta)^{1/2}} = \frac{k\Theta}{\tau_0 s}\left(\frac{2}{\pi}\right)^{1/2}\frac{1}{7{\cdot}5} = 0{\cdot}4\frac{\hbar n_1^{1/3}}{\tau_0} \qquad \ldots\ldots(4.4)$$

or

$$F_1 = 0{\cdot}4\frac{\hbar n_1^{1/3}}{m\mu(\theta)}. \qquad \ldots\ldots(4.5)$$

---

† The present paper deals with pure crystals only, but the effect of impurities can be taken into account in a similar way, as has been done previously (for a summary cf. Whitehead 1951).

F_1 is obtained in v cm^{-1} if μ is expressed in cm^2 v^{-1} sec^{-1}. If, for example, the density of ions is 10^{22} per cm^3, and if $\mu(\theta)=100$ then $F_1 \simeq 10^5$ v cm^{-1}, provided the effective mass is equal to the actual electronic mass.

For ionic crystals the breakdown strength according to (3.14) and (3.11) is ($\omega=(\epsilon/\epsilon_\infty)^{1/2}\omega_t$, $\omega_t/2\pi=$ frequency of residual rays)

$$F_c \simeq 0\cdot5\,\frac{me^2\omega}{\hbar}\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon}\right) \quad \text{if} \quad T_0 \ll \frac{\hbar\omega}{k}$$

$$eF_c \simeq 0\cdot8\,\frac{me^2\omega}{\hbar}\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon}\right) \quad \text{if} \quad T_0 \to \infty. \qquad \ldots\ldots(4.6)$$

The result is similar to that obtained by Callen (1949) but about 40% lower. To show the reason for this difference it should be remembered that the collision probability between electrons and the lattice vibrations is a function of electronic energy E with a maximum for a certain electronic energy. Hence an electric field $F(E)$ can be defined, such that the average rate of energy transfer from the field to electrons with energy E is equal to the rate of energy transfer from these electrons to the lattice vibrations. Like the above mentioned collision probability, this $F(E)$ has a maximum at an energy E_0 and Callen assumes $F_c=F(E_0)$. In the present treatment, owing to the required interchange of energy between electrons, it is the energy balance of all electrons that matters, and not that of electrons with a given energy.

Expression (4.6) can also be written in terms of the mobility using (4.1) with (cf. Fröhlich 1954, eqn. (7.18))

$$\frac{1}{\tau(T_0)}=\frac{e^2}{\hbar}\left(\frac{2m}{\hbar\omega}\right)^{1/2}\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon}\right)\bar{n}(T_0) \quad \text{if} \quad kT_0 \ll \hbar\omega. \qquad \ldots\ldots(4.7)$$

One then finds with (4.6)

$$F_c \simeq 0\cdot5\left(\frac{\hbar\omega}{2m}\right)^{1/2}\frac{1}{\mu(T_0)\bar{n}(T_0)} \quad \text{if} \quad kT_0 \ll \hbar\omega. \qquad \ldots\ldots(4.8)$$

Redfield (1953) found experimentally $\mu=250\pm50$ cm^2 v^{-1} sec^{-1} at $T_0=82^\circ$K on NaCl. Using $\hbar\omega/k=340^\circ$ and putting the effective mass m equal to the electronic mass, one has $F_c \simeq \frac{1}{2}\times10^6$ v cm^{-1}. This compares favourably with $0\cdot8\times10^6$ v cm^{-1} which according to R. Cooper (personal communication) is the most likely experimental value. It must be emphasized in this connection that, as Cooper and co-workers have shown (e.g. Calderwood, Cooper and Wallace 1953), the electric strength is very sensitive not only to impurities but also to crystal imperfections, which can be removed by careful annealing. This treatment reduces the electric strength which, theoretically, is understandable in terms of the scattering of electrons by imperfections.

Equation (4.6) is theoretically not very reliable because it involves the use of the perturbation theory which, for alkali halides for example, can not be justified as discussed in detail by Fröhlich (1954). Equations (4.3) and (4.5) for non-polar crystals, however, should be theoretically reliable because the interaction between electrons and lattice displacement is, in this case, much weaker than for ionic crystals. Together with (4.3) it offers for the first time a possibility of calculating the breakdown strength in terms of measurable quantities (ionic density n_i, mobility μ, Debye temperature Θ and effective mass m).

REFERENCES

BLOCH, F., 1928, *Z. Phys.*, **52,** 555.

CALDERWOOD, T. H., COOPER, R., WALLACE, A. A., 1953, *Proc. Instn Elect. Engrs*, Pt. II, **100,** 105.

CALLEN, M. B., 1949, *Phys. Rev.*, **76,** 1394.

FRÖHLICH, H., 1937, *Proc. Roy. Soc.* A, **160,** 280; 1947a, *Ibid.*, **188,** 521; 1947 b, *Ibid.*, **188,** 532; 1954, *Advances in Physics, Phil. Mag. Suppl.*, **3,** 325.

FRÖHLICH, H., and SEITZ, F., 1950, *Phys. Rev.*, **79,** 526.

KAWAMURA, H., OHKURA, H., KIKUCHI, T., 1954, *J. Phys. Soc. Japan*, **9,** 541.

O'DWYER, J. J., 1954, *Austr. J. Phys.*, **7,** 36.

PARANJAPE, B. V., 1953, *Report of the British Electrical and Allied Industries Research Association*, L/T 285; 1954, *Ibid.*, L/T 307.

PINES, D., 1953, *Phys. Rev.*, **92,** 626.

REDFIELD, A., 1953, *Phys. Rev.*, **91,** 753.

ROSE, A., 1955, *Phys. Rev.*, **97,** 1538.

SEITZ, F., 1948, *Phys. Rev.*, **73,** 549 ; 1949, *Ibid.*, **76,** 1376.

WHITEHEAD, S., 1951, *Dielectric Breakdown in Solids* (Oxford : Clarendon Press).

The Breakdown of Gases subject to Crossed Electric Fields

By W. A. PROWSE and P. E. LANE

Department of Physics, Durham Colleges in the University of Durham

MS. received 13*th August* 1954 *and in amended form* 5*th September* 1955; *read before the Physical Society Conference on Ionization in Gases held at Oxford, July* 1953

Abstract. The hypothesis is made that in ultra-high-frequency breakdown where the drift movement imposed on an electron by the field is very restricted, the localization of the electron contributes to breakdown. An experimental system is described in which an auxiliary unidirectional or relatively low-frequency field is imposed at right angles to a microwave field to destroy the localization. When the auxiliary field is unidirectional, or of relatively low frequency (0·86 Mc/s) its application raises the breakdown stress, but when the frequency is raised to 9·7 Mc/s the two fields appear to act quite independently in producing breakdown. A partial explanation is advanced. The gases studied include air, oxygen, nitrogen, hydrogen and neon.

§ 1. Introduction

Ultra-high-frequency breakdown is conveniently defined by the condition that even at breakdown stress an electron in the test gap has an amplitude of movement shorter than the gap length, i.e. electrons can survive in the gap without being captured by the electrodes. It is evident that electron multiplication can proceed by collision ionization in such circumstances without any need for a secondary process. Whether or not a given gap will break down under a given stress will depend on whether or not an indefinitely continued growth of electron population can occur. This may be limited by recombination, diffusion or attachment, but attachment need only be considered in electronegative gases and any general theory of breakdown must be based on the balance between collision ionization on the one hand and recombination or diffusion loss on the other.

In the most highly developed theory of ultra-high-frequency breakdown the general picture is that there exists a particular equilibrium stress E_s at which generation of new electrons by collision ionization just balances the removal of electrons out of the test gap by diffusion (Herlin and Brown 1948, Brown and McDonald 1949). If the applied stress E just exceeds E_s the gap will eventually break down. The greater the value of $E-E_s$ the less will be the time taken for the establishment of any given electron population in the gap, i.e. the shorter will be the formative time lag between the application of the stress and the breakdown of the gap. If the stress is applied for a limited time t shorter than that taken by an electron in diffusing out of the gap, the breakdown stress will increase progressively as t decreases.

This situation was observed (Labrum 1947) in the breakdown of neon by pulses at microwave frequencies: the results agreed with the hypothesis that electrons were not removed during a short pulse. Oscillographic records of the breakdown of neon by individual pulses at microwave frequencies (Prowse and Jasinski 1952) are in quantitative agreement with Labrum's hypothesis. The breakdown stress of neon in sustained ultra-high-frequency fields, measured by Gill and von Engel (1949) was always less than that measured in pulsed fields, for the same gas pressure.

Observations on polyatomic gases, however, have shown an entirely different pattern. In measurements on air, pulses at microwave frequencies gave the same values of breakdown stress as sustained oscillations in the hundred-megacycle region (Cooper 1947, Pim 1949). But air is an attaching gas and the results might be affected by this. Oscillograms of microwave breakdown in hydrogen and nitrogen have been recorded in sufficient number to make possible a statistical analysis which showed a normal statistical lag but a formative lag shorter than the resolving time of the apparatus (5×10^{-8} sec). The time intervals employed were in the range 0 to 2·5 microseconds, i.e. over this time interval the breakdown stress was independent of pulse duration (Prowse and Jasinski 1952).

It is thus implicit that at microwave frequencies the processes determining whether or not breakdown can take place in polyatomic gases are operative in times short compared with one microsecond. Moreover the oscillograms showed that the breakdown is exceedingly abrupt. It appears that at a certain stress there is a sudden transition to complete instability, not preceded by the gradual accumulation of a large electron population.

A distinctive feature of pre-breakdown conditions in ultra-high-frequency fields is the localization of the group of electrons whose multiplication eventually constitutes breakdown, for example in nitrogen at 760 mm Hg, at a frequency of 2800 Mc/s the amplitude of oscillation of an electron for a field just capable of producing breakdown is $1{\cdot}6 \times 10^{-3}$ cm. Diffusion contributes a comparable movement and a concentrated group of electrons would spread into a nebulous spherical cloud of radius 0·06 cm during a one-microsecond pulse (Prowse and Jasinski 1951). Even at breakdown stress each electron makes many more exciting than ionizing collisions, particularly in polyatomic gases, so that the group of electrons moves not in a pure gas but in a gas containing excited molecules. Independent evidence exists that the presence of excited molecules facilitates breakdown (J. J. Thomson 1928).

An electron in a microwave field, apart from producing ionization, delivers energy to the gas and if this process contributes to breakdown an explanation of the onset of instability is evident—as more electrons are produced by collision ionization, so the rate of supply of energy to the gas in the neighbourhood is also increased, with a corresponding increase in the rate of collision ionization. The reversion of excited states is one of the few processes rapid enough to fit in with a formative time less than 10^{-7} second.

These suggestions are not intended to do more than indicate a possible physical ground to the specific hypothesis that the localization of the initial electrons may facilitate breakdown at ultra-high frequencies.

Applying a unidirectional electric field at right angles to the microwave field would cause the electrons to drift through the gas and on the view stated this would increase the microwave stress required to produce breakdown. This paper is concerned with an attempt to test this possibility (Prowse and Lane 1953).

§ 2. EXPERIMENTAL

The microwave field which is regarded as the main field was produced in a cylindrical resonator operating at 10 000 Mc/s in the H_{011} mode, in which the electric component takes the form of a closed ring (the E ring), the magnitude of the field falling off towards the boundaries of the resonator and towards the centre. In this mode the current flow in the resonator walls is entirely circumferential so that it is possible to insulate the flat ends from the curved cylindrical surface without materially impeding the oscillations (figure 1).

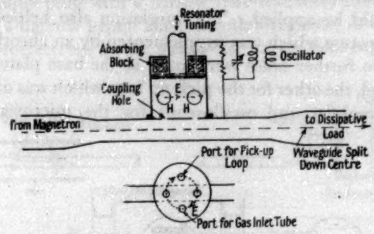


Figure 1. The resonator and tapered section of waveguide, showing the coupling holes, the method of applying auxiliary fields and the form of the microwave field.

The auxiliary field can be generated by applying suitable direct or oscillatory fields to the ends of the resonator: the cylinder itself can be regarded as a third electrode contributing to the uniformity of the field pattern. The relation between the field at any point and the applied potentials was determined by means of electrolytic tank tests on a model of the resonator.

A pulse-operated magnetron was employed to give one-microsecond pulses of oscillation frequency 10 000 Mc/s usually at a repetition rate of about 400 per second. In order to control the input of energy to the resonator a section of waveguide was equipped with a pair of tuning stubs and obstacles (figure 2). The combination produced standing waves of adjustable amplitude in the section of guide between the stubs. Coupling to the resonator was effected by means of two holes in the waveguide wall and end-plate of the resonator, a method giving preferential excitation of the desired mode (Bleaney, Loubser and Penrose 1947). In this mode the magnetic field at the end-plate of the resonator is radial and this was readily examined using a magnetic pick-up loop.

Apart from mode purity and from amplitude control, a chief requirement of the feed system was the production of a strong field in the resonator. This was limited chiefly by the total energy input and by the damping produced by the introduction of the glass vessel used as a gas container. To reduce this damping it was desirable to have a resonator neither too narrow nor too flat, and to achieve optimum energy transfer it was desirable that the half-wavelength in the guide $\frac{1}{2}\lambda_g$ should be equal to the diameter of the E ring in the resonator. λ_g could be increased by reducing the longer guide dimension, and this in turn would increase the longitudinal component of the magnetic field, which was the actual component feeding the resonator.

Calculations of sets of resonator and waveguide dimensions were made and a design was chosen giving a resonator of diameter 4·72 cm fed from a rectangular waveguide of longer dimension 1·90 cm. The feeding section of waveguide was made in the form of two tapered channels, fitting together so that the join lay in the zero-current plane (Labrum 1947). This design gave access to the coupling holes, which had to be adjusted for critical coupling by step-by-step enlargement.

The main constructional features of the resonator are shown in figure 1, from which it will be seen that the base (attached to the waveguide), the cylindrical wall and the adjustable outer end-plate were insulated from one another so that auxiliary fields could be applied. The insulation also helped to suppress unwanted modes, a process which was supplemented by an absorbing block behind the end-plate. Two further holes were made in the base plate, one for the inlet tube to the gas vessel, the other for the pick-up loop which was used in conjunction with a rectifier, amplifier and oscilloscope for the microwave measurements (figure 2).

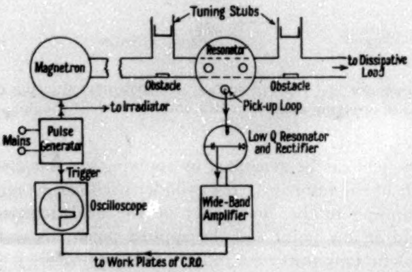


Figure 2. Block diagram of the electrical system, including the means adopted for producing standing waves in the section of guide feeding the resonator.

The gas chamber and associated equipment could be pumped out hard by means of an oil diffusion pump and cold trap, but baking out was not possible because of the location of the gas vessel inside the resonator. Mercury vapour was completely excluded from the system and all pressure measurements were made by means of reflecting differential gauges.

The arrangements for display and measurement are shown in figure 2. An output corresponding to the envelope of the magnetic field in the resonator was obtained from the crystal rectifier and was amplified for application to the oscilloscope. Provision was made in the pulse-forming circuits for generating a pulse to trigger the time base. Relative amplitude measurements were made by applying backing-off voltages to the work plates.

After some experience with this apparatus it became evident that frequent standardization would be desirable. Provision was accordingly made for supplying electrolytically generated hydrogen to the gas vessel so that all values of breakdown stress could be referred to that of pure dry hydrogen at a standard pressure.

In order to reduce the long statistical lag associated with the wait for a casual electron to appear in the strong field region, radiation from an auxiliary spark was passed into the test chamber immediately before the onset of the pulse. Some details are shown in figure 3, where it will be seen that the spark gap was connected between two cylindrical liners in a long tube aligned with the gas inlet. Capacitative connection was made to these by means of a pair of sleeves surrounding the outside of the tube. A high-voltage pulse applied to the sleeves produced an impulsive spark in the gap. The distance between the spark and the test chamber could be adjusted with a magnet. At the test chamber end of the tube an opaque plug with a narrow hole restricted the radiation from the auxiliary spark to a narrow pencil aligned with a segment of the E ring, so as to provide some photoelectrons in the gas in the strong field region.

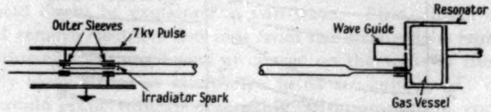


Figure 3. The irradiator system.

The efficacy of this method of irradiation had been established in earlier experiments (Prowse and Jasinski 1951) both for microwave breakdown and for unidirectional pulses. It has the advantage of providing electrons in the body of the gas and it has been shown that the breakdown streamers characteristic of microwave breakdown at moderate pressures develop from the track of the irradiating beam. It would seem at first sight that this form of irradiation simply provides a few casual electrons conveniently placed to avoid electrode effects, but in fact when the irradiating spark is within a few centimetres of the test vessel a slight lowering of breakdown potential occurs. This has been the subject of a separate study (Prowse and Lane 1956), but in the present experiments the distance has been kept fixed at 19 cm, at which distance the breakdown stresses agree closely with those observed using a radio-thorium source to provide the casual electrons. Without irradiation the statistical lags were so long as to make observations impracticable, except perhaps in hydrogen.

§ 3. OBSERVATIONS USING A UNIDIRECTIONAL AUXILIARY FIELD

A sustained unidirectional field would tend to produce a sheath of ions on the inner surfaces of the end walls of the gas vessel, so that the magnitude of the field applied to the gas in the region of the microwave E ring would be uncertain, particularly after the first discharge. In order to reduce this effect a separate auxiliary unidirectional pulse was produced for application to the end plates of the resonator, and the pulse-forming circuit for the magnetron was modified so that the beginning of the auxiliary pulse initiated the magnetron pulse. A potential divider network was used to control the fraction of the auxiliary pulse applied to the resonator end-plates.

3.1. *Dry Air*

In general the application of the auxiliary field caused an increase in the microwave onset stress for breakdown. Above a certain value of auxiliary field the character of the breakdown showed a change. When breakdown was produced by the microwave field alone, it was shown on the oscilloscope as a sudden collapse of the pulse envelope and every subsequent pulse in a train showed the same collapse. But in the presence of a sufficiently strong auxiliary field a second type of breakdown occurred in which only an occasional group of pulses in a train showed collapse, giving a flickering effect on the screen. The two types, referred to for convenience as A (absolute) and F (flicker), appeared at sharply defined amplitudes as separately shown on the graphs for hydrogen (figure 4). In breakdown of the A type the discharge continued when the irradiating spark was switched off, but in F type breakdown it did not.

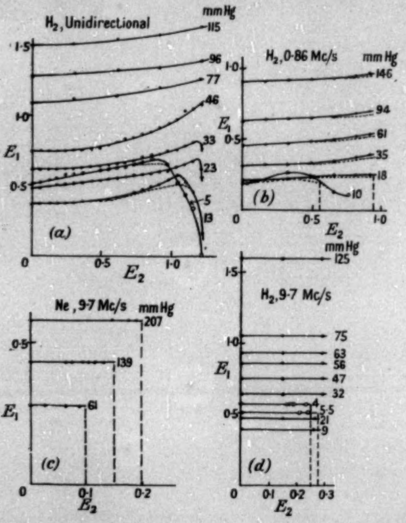


Figure 4. Typical graphs showing breakdown stresses for crossed fields of various frequencies. The microwave stress is plotted vertically in kv cm^{-1} (peak). Broken lines correspond to F (flicker) type breakdown.

3.2. *Hydrogen*

The hydrogen used in most of the experiments was obtained from the electrolysis of barium hydroxide solution: it was stored over phosphorus pentoxide before use. Although repeated tests were made, no difference was found in breakdown stress between this electrolytic hydrogen and commercial cylinder hydrogen. It is thought that gas purity is of less importance in pulse work than in work with sustained fields because of the relatively small number of collisions made by a given particle during a pulse. The results are shown in figure 4 where it will be observed that an increase of microwave breakdown stress generally occurs when the auxiliary field is applied. At the lower pressures the sharp dip for the highest auxiliary stresses may simply indicate an approach to the region at which the auxiliary pulses alone would produce breakdown. Two types of breakdown were again observed.

3.3. *Other Gases*

Generally similar experiments were made using oxygen and neon. In both these gases an increase in the microwave breakdown stress was observed when the auxiliary field was applied. With neon it was possible at some pressures to trace the complete curve to the point at which the auxiliary field alone would initiate breakdown. The two types of breakdown were manifested in oxygen, but in neon only the A type was observed.

Although the results of the experiments described tend to support the hypothesis they were designed to test, more than one interpretation could be offered. Starting from the assumption that a critical minimum number of electrons must be produced before breakdown becomes apparent, the inhibiting effect of the auxiliary field might be explained in two ways. First, direct clean-up to the walls could remove electrons and ions from the strong field region in the early stages, producing screening layers of charge on the walls of the vessel. (This is not really likely so far as microwave fields are concerned: detuning of the resonator would occur with an observable fall in amplitude, contrary to observation). Second, a uniform electron population drifting through a localized strong field region will not show the same increase by collision as a non-drifting population.

Calculations based on considerations given later (§6) showed that over the greater part of the experimental range, in all of the gases used, an electron would drift a distance comparable with the length of the gas vessel under the action of the auxiliary field for one microsecond. The oscillation amplitude produced by the microwave field would not exceed a fraction of a millimetre.

In all of the measurements sufficient time was allowed between observations for the vessel to become clear of products of previous discharges: this was checked by frequent refilling with fresh gas. It may be noted that breakdown stresses could be observed either by means of step-by-step increases in field strength, with the irradiating spark on, or by starting with the irradiator off at a field strength above breakdown value and seeing whether applying the irradiator started a discharge, thereafter bringing the field strength down in steps with suitable pauses between adjustments. The two methods gave identical values, but the first was quicker and was generally used.

§ 4. OSCILLATORY AUXILIARY FIELD

Let the microwave field be represented by ${}_0E_1 \sin \omega_1 t$ and the auxiliary field by ${}_0E_2 \sin \omega_2 t$, where ${}_0E_1$ and ${}_0E_2$ are complex. Then the instantaneous value of the field is $({}_0E_1{}^2 \sin^2 \omega_1 t + {}_0E_2{}^2 \sin^2 \omega_2 t)^{1/2}$. But $\omega_1 \gg \omega_2$ and the random velocity of the electron will not follow instantaneous values of the microwave field (Compton 1923). In the experiments to be described the frequency of the auxiliary field did not exceed 10 Mc/s, which is quite slow enough for the random velocity to follow the pattern of the oscillations, so that over any short interval of time the effective value of the field is $[{}_0E_2{}^2 \sin^2 \omega_2 t + ({}_0E_1{}^2 \sin^2 \omega_1 t)_{\mathrm{av}}]^{1/2}$ where 'av' indicates the mean value over a microwave cycle. Thus the value of the field deciding the random velocity is $(E_1{}^2 + {}_0E_2{}^2 \sin^2 \omega_2 t)^{1/2}$, where E_1 refers to the r.m.s. value of the microwave field.

To apply an oscillatory auxiliary field, a simple power oscillator was loosely coupled to a resonant circuit, in which the end plates of the resonator were placed in parallel with the tuning condenser. A twin diode voltmeter was used to measure the voltage developed across the end plates of the resonator (Gill and von Engel 1949) and the actual field was calculated with the aid of information given by the electrolytic tank experiments. Frequencies of 0·86 Mc/s, 2·66 Mc/s and 9·7 Mc/s were employed.

It was not feasible to apply fields of this type as pulses coinciding with the magnetron pulse: at the lowest frequency the pulse duration was less than that of one complete cycle. The pulse repetition frequency was fixed by a relaxation oscillator and so would not remain constant enough for each pulse to select the same part of the wave form of the auxiliary field: the various pulses could be considered to provide all possible combinations of the auxiliary field and the microwave field. Since these fields were at right angles in space the combined field would always be greater than the microwave field acting alone, except for the instants when the auxiliary field passed through its zero value. These instants are of importance: momentarily only the microwave field is acting, the formative time for microwave breakdown is very short in polyatomic gases, so there is always a possibility of breakdown at that value of the microwave field which acting alone would produce breakdown. This would mask any increase in microwave breakdown stress caused only by the drift of electrons in the auxiliary field.

In the figures showing the results of the crossed-field experiments the microwave stress is plotted vertically and the auxiliary stress horizontally. Thus if breakdown were determined by the combined fields only, the points for a given pressure would lie on an elliptical arc joining the points on the axes, i.e. those for which breakdown is produced by the microwave field alone in one case, and by the auxiliary field alone in the other. It may be noted that the maximum value of the combined fields for any point on the graph is given by the length of the line joining that point to the origin.

4.1. *Auxiliary Field Frequency* 0·86 *Mc/s* (*Hydrogen and Neon*)

In hydrogen the application of the auxiliary field produced an increase in the microwave breakdown stress (figure 4 (*b*)) the increase continuing over the range of auxiliary voltage available except at pressures below some 20 mm Hg, where the graphs tended to turn over as they approached the region where the auxiliary field alone would produce breakdown. The discharge produced by the auxiliary field was weak in character and did not noticeably damp the microwave oscillations in the resonator. When the microwave amplitude was raised in the presence of this discharge a sudden bright glow appeared with collapse of the oscilloscope trace. On the graph the corresponding microwave stress appeared to lie on the continuation of the ordinary breakdown curve. This effect was observed at all frequencies. In the presence of the auxiliary field both the A and F types of breakdown were observed in hydrogen, but only the A type in neon.

In neon the application of the auxiliary field at the lower pressures produced a marked increase in the microwave breakdown stress to begin with, followed by a flat portion and a turn over as the auxiliary field breakdown was approached. The initial rise was missing at the higher pressures. If this absence were associated with the shorter electron ambits at the higher pressures, the use of a higher frequency auxiliary field might be expected to result in an extension of the simpler form of curve to lower pressures. Calculation of the electron ambits, by a method described below, showed that in the region of the initial rise these ambits were small compared with the length of the gas vessel.

4.2. *Auxiliary Field Frequency 2·66 Mc/s (Hydrogen, Neon, Nitrogen, Oxygen and Air)*

With all of these gases (figure 5) the same general feature was very noticeable, namely that the microwave breakdown stress was independent of the magnitude of the auxiliary field except at the lower pressures. Also in hydrogen, nitrogen and oxygen it was observed that at the lower pressures, where the microwave breakdown stress showed a slight increase with increasing auxiliary field, the two types of discharge again appeared.

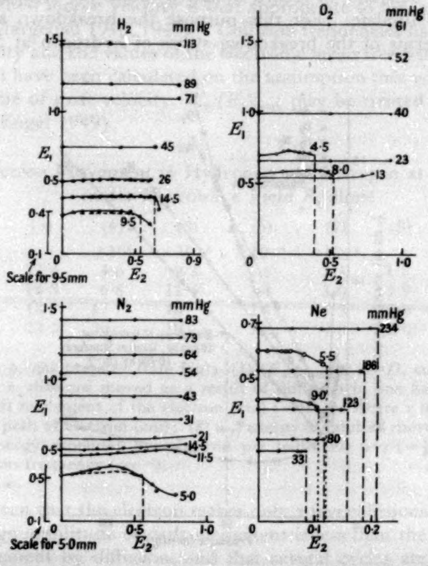


Figure 5. Typical graphs showing breakdown stresses for crossed fields. Auxiliary field frequency 2·66 Mc/s. Microwave stress plotted vertically in kv cm^{-1} (peak). Open circles are used for pressures below that for minimum breakdown stress. Dotted lines correspond to F type breakdown, vertical broken lines to breakdown by the auxiliary field alone.

In neon at all pressures it was possible to produce breakdown by the auxiliary field alone. The microwave field appeared to act independently of the auxiliary field right up to this boundary for pressures above that corresponding to minimum microwave breakdown stress (figures 5 and 6). Compared with the auxiliary field breakdown stress, the microwave breakdown stress in neon was relatively high. This was not unexpected as the ultra-high-frequency breakdown stress in neon is known to increase with decreasing pulse length.

4.3. *Auxiliary Field Frequency 9·7 Mc/s (Hydrogen, Neon and Air)*

At a frequency of 9·7 Mc/s, in all three gases examined, the microwave breakdown stress was independent of the auxiliary stress within experimental error (figures 4 (*c*), (*d*)).

§ 5. MICROWAVE BREAKDOWN STRESSES—NO AUXILIARY FIELD

Figure 6 shows microwave breakdown stress as a function of pressure, for zero auxiliary field, for several gases. Even when the microwave field alone is considered it cannot safely be assumed that there exists a true breakdown stress independent of electrode configuration (Herlin and Brown 1948), but for this very reason it is desirable to record tentative values for comparison with other data as they accumulate. For this purpose the breakdown stresses were all determined in terms of the breakdown stress of hydrogen at 50 mm pressure, given by earlier work as 790 v cm^{-1} (Prowse and Jasinski 1952). The pressure range of the present experiments only extends into the lower portion of the range of the earlier experiments. Taking the single value of breakdown stress in hydrogen to establish the scale, the new curve for hydrogen formed a linear continuation of the curve of the earlier experiments, but the new curves for other gases did not quite join the old; that is, fitting the old and the new hydrogen curves together for one point also fitted them for slope: hydrogen gave a

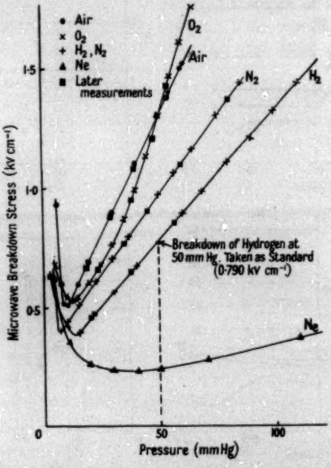


Figure 6. Microwave breakdown stresses (peak values) for various gases, no auxiliary field acting. The later measurements were made after the lapse of several months.

unique scale factor. This was not true of air, where a given scale factor caused the old and the new curves to intersect at a finite angle. Naturally the numerical results for microwave breakdown stresses determined in this way are to be accepted with more reserve than the relative values, with which this paper is chiefly concerned.

The measurements on the polyatomic gases were repeated after several months and the new values agreed with the old within the precision of observation. As in the earlier experiments, oxygen does not seem to conform to so simple a pattern as the other gases.

§ 6. DISCUSSION

When the microwave field E_1, alone, is under consideration, i.e. when there is no auxiliary field, approximate values for the quantities needed to describe the electron movement can be calculated from the data given by Townsend (1947), assuming that the random velocity is that appropriate to the r.m.s. value of the applied field (Margenau 1946, 1948). Collision frequencies have been calculated from this velocity and the values of the electronic mean free path. Amplitudes of drift movement have been calculated on the assumption that where W_0 (cm sec^{-1}) is the peak value of drift velocity, $W_0/(E_1)_{\text{peak}}$ may be treated as a true mobility (Gill and von Engel 1949).

Table 1. Electron Movement in Hydrogen and Nitrogen at Breakdown Stress under Microwave Field E_1 alone

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
			$\times 10^3$	$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$			$\times 10^{11}$
H_2	90	13·4	4·6	9·3	1·6	2·3	1·7	0·15	3·3
N_2	90	16·8	6·8	11·4	2·4	3·0	1·9	0·28	2·7
H_2	30	18·7	17	18	2·1	7·3	2·0	0·09	1·2
N_2	30	23·3	20	19	3·0	9·0	2·0	0·16	0·9

(1) Gas; (2) p, gas pressure (mm Hg); (3) $(E_1/p)_{\text{peak}}$; (4) D, coefficient of diffusion (cm^2 sec^{-1}); (5) r, distance moved as a result of diffusion in one half-cycle (cm); (6) q, amplitude of drift movement of the electron (cm) ($=W_0/\pi\nu$ where ν is applied frequency); (7) λ, mean free path of electron (cm); (8) w_1, energy of random movement of the electron (ev); (9) w_2, energy expended per electron per half-cycle (ev) ($=\frac{1}{4}\pi q(E_1)_{\text{peak}}$); (10) n_2, electronic collision frequency (sec^{-1}).

It will be seen that the electron makes only a few collisions in each half-cycle, that the average amplitude of drift movement is less than the mean free path or than the movement by diffusion, and that several cycles are required to bring the average random velocity to its terminal value. Because of scattering, the chance that an electron will collide with a molecule which it has itself previously excited is small, assuming that the collision cross section of the excited molecule is of the same order as that of the neutral molecule.

The values of E_1/p corresponding to the minima in the curves of figure 6 have been used, in conjunction with Townsend's data, to calculate the electron collision frequencies for these minima, giving the following values: hydrogen $5{\cdot}6 \times 10^{10}$ sec^{-1}, nitrogen $2{\cdot}7 \times 10^{10}$ sec^{-1}, oxygen $4{\cdot}1 \times 10^{10}$ sec^{-1}, air $3{\cdot}5 \times 10^{10}$ sec^{-1}, neon $4{\cdot}1 \times 10^{10}$ sec^{-1}. These may be compared with the angular frequency of the applied field, namely $6{\cdot}3 \times 10^{10}$ sec^{-1}. It is recalled that there are theoretical grounds for expecting maximum ionizing efficiency when the applied angular frequency is equal to the electron collision frequency (Townsend and Gill 1938, Margenau 1946).

When the auxiliary field E_2 is applied it causes electrons to drift to and fro parallel to the axis of the resonator. The amplitude of this movement in the absence of the microwave field is readily calculable from published data on electron drift velocities, assuming that the random velocity V and drift velocity W take values corresponding to the instantaneous values of E_2/p in the cycle. When E_1 is applied as well as E_2 it is to be expected that the random velocity of the electron will be determined by the combined field, viz. $(E_1^2 + {}_0E_2^2 \sin^2 \omega_2 t)^{1/2}$. This will somewhat reduce the drift velocity because, in the present experimental range, the drift velocity is inversely proportional to the random velocity (Townsend 1947). Let W' be the corrected value of the drift velocity.

Complete graphical integration of the (W', t) curve over a half cycle was tried and was compared with the result obtained by taking the value of W' corresponding to $(E_2/p)_{\text{peak}}$ and assuming a sinusoidal variation of drift velocity during the cycle. The difference was small enough for the latter method to be considered satisfactory.

It has already been suggested that when an auxiliary field is superposed on the microwave field, breakdown may be initiated when the auxiliary field is momentarily zero. This agrees with the observed condition that over most of the experimental range the microwave field is independent of the magnitude of the auxiliary field. But if the amplitude of axial movement of the electrons is greater than half the inside length of the gas vessel, those electrons which are in the strongest part of the microwave E ring when the auxiliary field is zero will strike the glass wall in the next half cycle and may be trapped there. This condition should correspond with the beginning of an increase of microwave breakdown stress with increasing auxiliary field. The observations in hydrogen at 0·86 Mc/s are not satisfactory for testing this possibility because the electron ambit exceeds half the capsule length over most of the experimental range, and at 9·7 Mc/s the ambit is always less than this critical length; it is seen however, that at 0·86 Mc/s the microwave breakdown stress increases with the auxiliary stress whereas at 9·7 Mc/s the microwave breakdown stress is entirely independent of the auxiliary stress.

The beginning of a rise of microwave breakdown stress with increasing auxiliary stress occurs as given in table 2.

The beginning of the rise is in no case sharply defined; for neon it occurs below the lowest value of the auxiliary field, viz. 33 volt cm^{-1}.

For the polyatomic gases the calculated electron ambit is sufficiently nearly equal to the half-length of the gas vessel, in agreement with the suggestion that breakdown can be initiated when E_2 is momentarily near zero provided that the electrons in the strongest part of the microwave field are not trapped by the walls in the next half -cycle.

In neon the formative time of breakdown is too great for such a mechanism to apply and the data of table 2 indicate that a rise of breakdown stress occurs for quite a small movement of the electrons at right angles to the microwave field. At 9·7 Mc/s the drift movement of the electrons is inappreciable, even for the strongest auxiliary field. At a pressure of 207 mm Hg the maximum value of E_2 used was 180 v cm^{-1}, giving a drift amplitude less than $3{\cdot}4 \times 10^{-2}$ cm (using an uncorrected value of W) and a diffusion movement in 1 μsec of $1{\cdot}7 \times 10^{-1}$ cm, so

Table 2. Values of Auxiliary Stress E_2' at which rise of Microwave Breakdown Stress begins. Inside length of gas vessel 2·2 cm

Gas	Freq. (Mc/s)	E_2' (v cm^{-1})	p (mm Hg)	Calc. elec. ambit (cm)
H_2	2·66	<500	9·5	<2·3
	2·66	<600	14·5	<1·9
N_2	2·66	130	5·0	0·62
	2·66	190	11·5	0·68
	2·66	240	14·5	0·71
	2·66	380	21·0	0·85
O_2	2·66	160	4·5	0·74
	2·66	150	5·8	0·80
	2·66	500	13·0	0·92
Air	2·66	155	5·0	1·07
	2·66	200	10·0	0·88
	2·66	300	15·0	0·90
	2·66	500	21·5	1·10
Ne	0·86	33	50	0·19
	0·86	33	98	0·12
	0·86	33	145	0·11
	0·86	33	190	0·08

that the movement caused by the field E_2 is much less than that caused by diffusion. This explains the absence of a rise of microwave breakdown stress, but it does not explain the apparent independence of action of E_1 and E_2 in producing breakdown.

The appearance of two separate forms of breakdown (A and F) in polyatomic gases occurs when the auxiliary field is great enough to produce a rise in microwave breakdown stress, and is probably caused by clean-up of ionization to the walls of the vessel between pulses. Thus the F type breakdown stops when the irradiating spark is stopped. It seems likely that at a somewhat higher value of the microwave field the space-charge fields at the walls of the vessel become strong enough to prevent further trapping, giving a transition to A type breakdown, which does not stop when the irradiating spark is switched off.

According to the views expressed, when the auxiliary field is applied (in polyatomic gases) breakdown is only likely to be initiated when the auxiliary field is instantaneously near zero, i.e. the effective duration of the pulse is much less than its actual duration. This was examined as follows. With zero auxiliary field, the microwave field was adjusted to the least value giving breakdown, next the microwave field was increased by a measured amount and the average time lag between switching on and breakdown was recorded. This was repeated until a curve could be plotted relating time lag and overvoltage. An auxiliary field of $900\,\text{v cm}^{-1}$ at 2·66 Mc/s was next applied and the process was repeated. It was found that at a given percentage overvoltage the time lag was much greater when the auxiliary field was applied, implying that the effective duration of the pulse had been reduced by the action of the auxiliary field.

ACKNOWLEDGMENTS

The authors wish especially to thank Dr. Elizabeth Laverick and Dr. W. Jasinski who set up and tested the first form of the apparatus and made the first tentative readings on which the subsequent development was based. They wish also to thank the British Electrical and Allied Industries Research Association for a grant towards the cost of the apparatus, and one of them (P.E.L.) gratefully acknowledges the receipt from the Department of Scientific and Industrial Research of a maintenance grant.

REFERENCES

BLEANEY, B., LOUBSER, J. H. N., and PENROSE, R. P., 1947, *Proc. Phys. Soc.*, **59,** 185.

BROWN, S. C., and MCDONALD, A. D., 1949, *Phys. Rev.*, **76,** 1629.

COMPTON, K. T., 1923, *Phys. Rev.*, **22,** 333, 432.

COOPER, R., 1947, *Proc. Instn Elect. Engrs*, Pt. III, **94,** 315.

GILL, E. W. B., and VON ENGEL, A., 1949, *Proc. Roy. Soc.* A, **197,** 112.

HERLIN, M. A., and BROWN, S. C., 1948, *Phys. Rev.*, **74,** 291.

LABRUM, N. R., 1947, *C.S.I.R., Australia*, RPR 85.

MARGENAU, H., 1946, *Phys. Rev.*, **69,** 508; 1948, *Ibid.*, **73,** 297.

PIM, J. A., 1949, *Proc. Instn Elect. Engrs*, Pt. III, **96,** 117.

PROWSE, W. A., and JASINSKI, W., 1951, *Instn Elect. Engrs*, Monograph No. 10; 1952, *Instn Elect. Engrs*, Monograph No. 32.

PROWSE, W. A., and LANE, P. E., 1953, *Nature, Lond.*, **172,** 116; 1956, *Appl. Sci. Res.* (in the press).

THOMSON, J. J., 1928, *Proc. Phys. Soc.*, **40,** 79.

TOWNSEND, Sir JOHN, 1947, *Electrons in Gases* (London : Hutchinson).

TOWNSEND, J., and GILL, E. W. B., 1938, *Phil. Mag.*, **26,** 290.

Fracture of Glass by Spherical Indenters

By J. P. A. TILLETT
Imperial Chemical Industries Limited, Akers Research Laboratories, Welwyn, Herts.

Communicated by H. Kolsky ; MS. received 25th July 1955

Abstract. The production of ring cracks in silicate glass, by the impact of steel balls and by the application of a static load to spherical indenters, has been studied for a large range of indenter curvature. It has been found that in impact tests a greater height of fall is required to produce ring cracks than predicted by a critical stress criterion of fracture. In static experiments, two different relations between the force required to produce fracture and indenter size have been found. Below a radius of curvature of approximately 3·5 cm the force is proportional to the radius and above this value to the square of the radius. Impact tests have also been carried out on a specimen of an organic glass-like material, with results similar to those of the impact experiments on silicate glass.

---

§ 1. Introduction

The formation of ring cracks on glass by the static pressure of a spherical indenter and by impact has been studied on many occasions and the results used in several cases to measure or define the strength and hardness of glass (Auerbach 1891, Raman 1926, Andrews 1930, Longchambon 1934, Bailey 1937). This paper reports further experimental work on the production of Hertzian ring cracks but the range of indenter radius covered was larger than in previous work and certain new results have been obtained. Experiments in which a hydraulically applied load has been slowly increased to the point of fracture (static measurements) and experiments involving the impact of freely falling steel balls have been performed on a single sample of silicate glass. Impact tests have also been carried out on a brittle organic glass (glycerol sextol phthalate).

If it is assumed that glass cracks when the tensile strength is exceeded at some point, the Hertz stress distribution for spherical indenters predicts the formation of circular cracks at the rim of the area of contact. The normal stress to produce such a tension may be calculated (see Timoshenko 1934 for a summary of the Hertz equations). However, it has been known for a long time from static measurements that with indenters of small radius of curvature, the tensile strength calculated from the observations in this manner is very much greater than the value obtained from measurements involving uniform tension (Auerbach 1891, Powell and Preston 1945). Griffith (1920) first put forward the idea of the existence of submicroscopic flaws in glass and it has been suggested by several authors that the various values obtained for the strength of glass, when measured under different conditions, may be due to the flaw distribution. Fisher and Holloman (1947) and Gibbs and Cutler (1951) have considered the value of the strength of glass on the basis of flaw statistics.

During work to investigate the use of impinging steel balls to measure the dynamic tensile strength of glass, it was found by the author that the height of fall to produce a ring crack varied systematically with the radius of the ball. It would at first sight appear normal that glass should break more readily under the impact of a large body but in this case, where the glass plate is rigid and bending cannot take place, the Hertz stress distribution predicts that the height should be independent of the ball radius. The actual relation that was obtained has been shown by Roesler (1956) to be equivalent to two other empirical laws governing the production of ring cracks found by Longchambon and Auerbach.

The growth of fracture was considered by Griffith to depend on the relative value of the strain energy density in the material and the energy required to form new surfaces. For spherical indenters, Roesler has shown that the empirical laws seem to agree with this basic law, and they are apparently due to the amount of strain energy available for the particular relation between force and area of contact in the Hertz stress field.

The results of the experiments described below indicate the limit of applicability of this energy criterion of fracture. They also show in what circumstances fracture by spherical indenters may be used to compare the strength and hardness of glasses.

§ 2. Experimental

The experiments on silicate glass were carried out using specimens cut from a sheet of 1 in. thick plate glass manufactured by Pilkington Bros. Ltd. With the largest indenter in the static experiments, the initial results indicated that the specimen was not sufficiently thick. A specimen made up of two 1 in. thick pieces of glass, cemented together with ' Araldite ' was therefore used instead. The surface of the glass was not specially treated but the specimens were handled as little as possible. The quoted strength of this glass at 1% risk was 1050 to 1500 lb for a sustained load and 2500 to 3000 lb for a momentary load.

2.1. *Impact Measurements*

The height of fall to produce ring cracks was measured for different sized ball bearings which ranged from 3/64 to 3/4 in. in radius. The ball was released from an electromagnet mounted above the specimen and the height of fall measured to the nearest millimetre. The mean value of the height to produce a ring crack was obtained and 5 to 10 determinations made for each ball size.

Subsequently, a very thin layer of grease was rubbed on to the ball and the size of the cracks that were produced when it fell from the critical height and above was measured by means of a travelling microscope. If the cracks were viewed by light reflected off the surface at an angle, the contact area was easily observed as a dark circle of grease and its diameter measured.

2.2. *Static Measurements*

The force between the indenter and glass specimen was increased slowly, using a hydraulic press, until a ring crack was produced and the value of the force at which fracture took place was measured for indenters ranging from $\frac{1}{8}$ in. to 5 in. in radius of curvature. The apparatus is shown in figure 1. The force was determined from the reading of the pressure gauge, which measured the oil pressure on the piston. In order to cover the large range of load in these experiments different pressure gauges were used. Between $\frac{1}{8}$ in. and 1 in. radius the indenters were steel balls but above this the indenters consisted of short cylinders of silver steel, 1 in. in diameter, having hardened, polished, spherical caps of known radii of curvature.

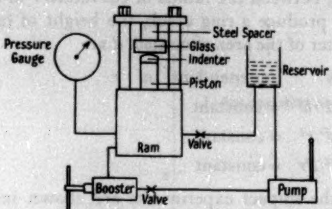


Figure 1. The hydraulic press used for applying a force between the spherical indenter and glass specimen.

The face of the glass in contact with the indenter was viewed through one side face and, when a pressure was applied, the area of contact was visible as a black disc. The force was increased until a ring crack was seen to start and run round the area of contact. The larger the indenter, the more violent was the initial cracking. Near the critical load, the force was applied at a constant rate of 7 lb sec^{-1} after an initial load, which depended on the indenter size, had been applied more rapidly. Measurements were made at half and double this rate of loading with no significant change in the mean value of the force required to cause cracking.

Because of the scatter in the results, especially with the indenters of large radius, each experiment was repeated at least twenty times. The diameters of the cracks were measured using a travelling microscope.

2.3. *Measurements on Glycerol Sextol Phthalate*

Glycerol sextol phthalate is an organic glass (a mixed glycerol–cyclohexanol ester of phthalic acid) which softens at about 50°c and at 20°c has a viscosity of 10^{11} poises and a Young's modulus of 2×10^{10} dyn cm^{-2}. Static measurements were not possible with this material because plastic flow rather than fracture took place as the pressure was increased. The specimen was cast in the form of a thick plate, 6 in. square and 2 in. thick. Dynamic measurements were possible for balls ranging from 3/32 in. to 1/2 in. in radius but outside this range ring cracks were no longer formed.

2.4. *Measurements on a Scratched Surface*

In order to examine the effect of surface scratches, the static measurements were repeated on silicate glass the surface of which was covered with a network of fine diamond scratches approximately 3 mm apart. The indenters were placed so that the circumference of the area of contact when fracture took place was within $\frac{1}{2}$ mm of a scratch. The force was increased until the flaw which developed along the diamond crack started to run in a circle around the contact area. In most cases an incomplete ring crack was formed.

§ 3. RESULTS

If a ring crack is produced when the tensile strength is exceeded at the circumference of the area of contact, the relations that would be expected from the Hertz equations, between the radius of curvature r of the indenter, the static force P required to produce a ring crack, the height of fall h to produce a ring crack and the diameter of the area of contact d are

Impact loading h independent of r ......(1)

$d/rh^{1/5} = \text{constant}$ ......(2)

Static loading $P/r^2 = \text{constant}$ ......(3)

$d^3/Pr = \text{constant}$ ......(4)

The results of the impact experiments are shown in figure 2, which is a logarithmic plot of the height of fall to produce fracture against the radius of the ball. The best line through the experimental points for silicate glass has a slope of $-1{\cdot}65$, suggesting a relationship of the form

$$hr^{5/3} = \text{constant} \qquad \text{......(5)}$$

within the range covered. The area of contact was equal to the value, corresponding to the critical height, calculated from the Hertz equations but there was a variation in the size of the cracks produced in similar circumstances. Sometimes the crack started on the edge of the area of contact; occasionally its diameter was 1½ times that of the area of contact, but in the great majority of cases the crack diameter was 12% greater than the contact diameter. The variation in size is probably due to the flaw distribution but the reason for the crack displacement is not understood. If the relation between h and r in (5) is combined with (2), and allowance is made for the fact that the crack diameter D has been found to be proportional to the contact diameter d, the resulting expression is $D/r^{2/3} = \text{constant}$. In fact, the slope of a logarithmic plot of the crack diameter against radius of the ball was found to be 0·62.

The results of the impact experiments on glycerol sextol phthlate are also shown in figure 2. With this material the best line through the experimental points has a slope of $-1{\cdot}90$ but the relation between crack diameter and ball size was the same as for silicate glass.

Figure 3 shows the results of the static experiments on silicate glass and the two different relations between the force to produce fracture and indenter radius which have been obtained. For small indenters, the force P to cause fracture is proportional to the radius of curvature of the indenter, as found by Auerbach, and as predicted by a critical energy criterion of fracture. However, for indenters larger than 3·5 cm in radius of curvature, the relation is more nearly $P/r^2 = \text{constant}$ as required by the critical stress criterion of fracture. The two lines which have been drawn through the experimental points have the theoretical slopes of 1 for small radii, and 2 for large radii of curvature.

In the static fracture, too, the cracks form slightly outside the area of contact and the crack diameter is approximately 16% greater than the diameter of the contact circle, as calculated from the equations of Hertz. Since the crack diameter is proportional to the contact diameter, if P/r^2 is constant D should be directly proportional to r, and if P/r is constant D should be proportional to $r^{2/3}$. The crack sizes obtained from the static experiments are shown in figure 4.

The lines through the experimental points have been drawn with a slope of $\frac{2}{3}$ for the small and 1 for the large indenters. It was possible to measure the size of the cracks produced by balls less than $\frac{1}{8}$ in. in radius although the force to produce such cracks was too small to be measured with the apparatus described.

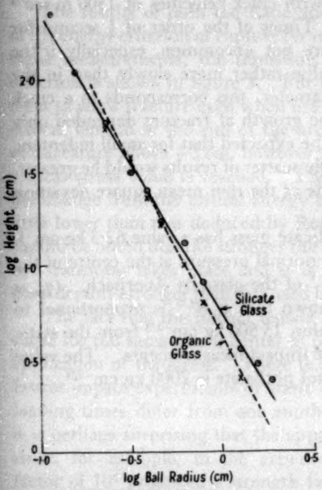


Figure 2. The effect of ball size on the height of fall required to produce a ring crack.

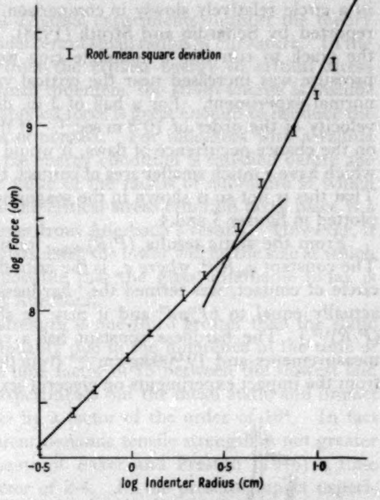


Figure 3. The effect of indenter curvature on the force required to produce a ring crack.

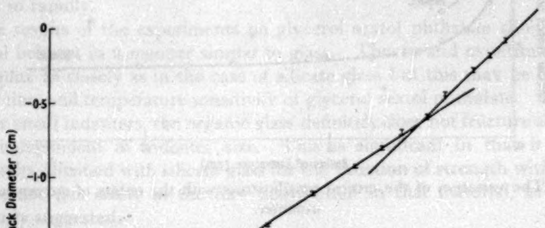


Figure 4. The variation of the crack diameter with the radius of curvature of the indenter.

The suddenness of the appearance of the cracks and the size of the cones produced increased with the radius of curvature of the indenter. For example, with a $\frac{1}{4}$ in. radius indenter only a faint ring was formed, whereas with one of 3 in. radius a cone rapidly extended to a depth of approximately 1 mm into the glass and the crack tended to become a spiral. For small indenters, the crack, which could be seen to start at one point just outside the contact area, ran round in a circle relatively slowly in comparison with crack velocities of 1500 m sec^{-1} reported by Schardin and Struth (1938). Times of the order of $\frac{1}{2}$ second for the crack to run into a complete ring were not uncommon, especially if the pressure was increased near the critical value rather more slowly than in the normal experiment. For a ball of 1 in. diameter, this corresponds to a crack velocity of the order of 10^{-3} m sec^{-1}. If the growth of fracture depended only on the chance occurrence of flaws, it would be expected that for small indenters, which have a much smaller area of contact, the scatter of results would be greater. That this is not so is shown in the magnitude of the root mean square deviation plotted in figures 3 and 4.

From the static results, $(P/R)_{\text{critical}}$ for silicate glass has a value 62·3 kg cm^{-1}. The constant $q_m R^{1/3}$, where q_m is the critical normal pressure at the centre of the circle of contact, was termed the ‘ hardness ’ of the glass by Auerbach. (q_m is actually equal to $6P/\pi d^2$ and it may be shown that $q_m R^{1/3}$ is proportional to $(P/R)^{1/3}$.) The hardness constant has a value 15 500 kg cm$^{-5/3}$ from the static measurements and 19 900 kg cm$^{-5/3}$ from the impact measurements. The value from the impact experiments on glycerol sextol phthalate is 2000 kg cm$^{-5/3}$.

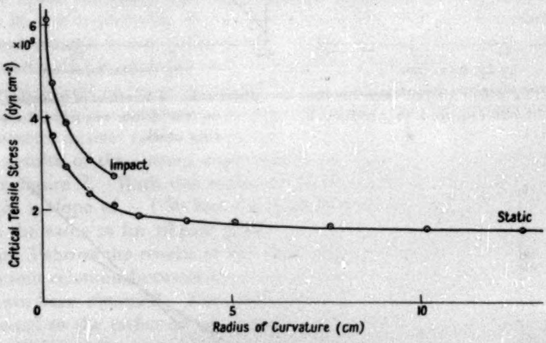


Figure 5. The variation of the critical tensile stress with the radius of curvature of the indenter.

In figure 5 is shown the variation of the critical tensile stress with indenter radius. This stress has been calculated, using the Hertz equations, from the value of the applied force when the glass breaks and is in fact equal to $q_m(1-2\nu)/3$ where ν is Poisson's ratio. As can be seen also in figure 3, the experimental points for the largest indenters do not fall exactly on the lines which indicate a constant critical stress for large radii of curvature.

When the surface was scratched with a diamond, there was again a difference in the results for small and large indenters. For indenters of radius 1 in. and less, the force to produce fracture was the same as for unscratched specimens but for larger indenters the force required was less. This reduction increased with indenter size and was, for example, 30% with a 4 in. radius indenter. The results, in fact, fell between the values required by the critical stress and critical energy criteria of fracture.

§ 4. DISCUSSION

The results of both the static and dynamic experiments support the theory of the energy criterion of rupture for indenters of small radius of curvature. In the static measurements, the transition from the critical energy to critical stress criterion is shown in figure 3. For small indenters, the strain energy is insufficient to allow fracture even when the applied force is great enough to produce the critical tension at the rim of the area of contact. For indenters with a radius of curvature above 3·5 cm, however, the energy condition is satisfied before the critical stress value is reached. The value of the radius of curvature at which the change from the critical energy to the critical stress criterion takes place is a little lower than that deduced by Roesler from Auerbach's results. However, it should be noted that the stronger the glass used, the lower will be the size at which the transition takes place and it is possible that the glass obtainable today is considerably stronger than that used by Auerbach.

The fact that the impact tensile strength is one-third greater than the static value for the same size indenter is probably due to the difference in the time of application of the stress. There is a time factor of 15 between the slowest and fastest impact experiments or static experiments but the mean static and impact loading times differ from one another by a factor of the order of 10^6. In fact it is perhaps surprising that the apparent dynamic tensile strength is not greater since, for example, in the experiments of Baker and Preston (1946) a time factor of 10^5 produced a strength factor of 2·4. In the present impact experiments, if a ring crack is to form during the time that an $\frac{1}{8}$ in. diameter ball is in contact with the specimen, the crack must run in the circle with a velocity of 140 m sec^{-1}. It is to be expected that a larger stress would be needed to produce a crack so rapidly.

The results of the experiments on glycerol sextol phthalate show that this material behaves in a manner similar to glass. Theory and experiment do not agree quite as closely as in the case of silicate glass but this may be due to the greater time and temperature sensitivity of glycerol sextol phthalate. Nevertheless, for small indenters, the organic glass definitely does not fracture at a critical stress independent of indenter size. This is significant in that it indicates that results obtained with silicate glass for the variation of strength with indenter size are not due solely to the flaw distribution in that material, as has been previously suggested.

Thus it may be seen that if indenters are used of such size that the force to produce a ring crack is proportional to the radius of curvature, then the production of this type of fracture provides a convenient method of obtaining results to compare the hardness of glasses.

ACKNOWLEDGMENTS

The author wishes to express her thanks to Dr. H. Kolsky and Dr. F. C. Roesler for advice, criticism and discussion.

REFERENCES

ANDREWS, J. P., 1930, *Proc. Phys. Soc.*, **43**, 18.
AUERBACH, E., 1891, *Ann. Phys. Chem.*, **43**, 61.
BAILEY, J., 1937, *J. Amer. Ceram. Soc.*, **20**, 42.
BAKER, T. C., and PRESTON, F. W., 1946, *J. Appl. Phys.*, **17**, 170.
FISHER, J. C., and HOLLOMAN, J. H., 1947, *Metals Technology*, **14**, T.P. 2218.
GIBBS, P., and CUTLER, I. B., 1951, *J. Amer. Ceram. Soc.*, **34**, 200.
GRIFFITH, A. A., 1920, *Phil. Trans. Roy. Soc.*, **221**, 163.
LONGCHAMBON, L., 1934, *C. R. Acad. Sci., Paris*, **199**, 1381.
POWELL, H. E., and PRESTON, F. W., 1945, *J. Amer. Ceram. Soc.*, **28**, 145.
RAMAN, C. V., 1926, *J. Opt. Soc. Amer.*, **12**, 387.
ROESLER, F. C., 1956, *Proc. Phys. Soc.* B, **69**, 55.
SCHARDIN, H., and STRUTH, W., 1938, *Glastechn. Ber.*, **16**, 219.
TIMOSHENKO, S., 1934, *Theory of Elasticity* (New York : McGraw-Hill).

Indentation Hardness of Glass as an Energy Scaling Law

By F. C. ROESLER
Imperial Chemical Industries Limited, Akers Research Laboratories, Welwyn, Herts.

Communicated by H. Kolsky; MS. received 25th July 1955

Abstract. A number of independently found empirical laws for the occurrence of Hertzian ring cracks are all equivalent and can be put in the form of an energy scaling law. It is suggested that this is the true explanation of the observed size effects, which would then lie in the dynamical similarity relations of geometrically similar stress fields and fracture surfaces.

---

§ 1. Auerbach's Law

Under the pressure of a spherical indenter, applied either quasi-statically or by impact, glass breaks first by a ring crack encircling the area of contact, mostly at some small distance from its boundary. The fractures caused by indenters of varying radius are geometrically similar. Very extensive experimental work has been done, by many investigators from Hertz (1882) to Tolansky and Howes (1954), on the generation of such cracks, and it has long been recognized that the condition for cracking is not that of a constant critical stress. There is a size effect, in the sense that the smaller the indenter is, the higher is the critical stress. Some empirical relations between the indenter radius and quantities which measure the critical stress, directly or indirectly, are listed below. P is the total force acting between indenter and specimen, r is the indenter radius, D is the diameter of the crack, q_m is the critical normal pressure at the centre of the circle of contact, H is a constant (Auerbach hardness), A is the area enclosed by the crack, i.e. $A = D^2\pi/4$, W is the kinetic energy of the indenter (steel ball) in an impact experiment, h is the height of free fall of the indenter (steel ball) in an impact experiment. The type of experiment and the author are given for each relation.

$$(P/r)_{\text{critical}} = \text{constant} \qquad \ldots\ldots(1)$$

(Auerbach 1891, quasi-static ;
Tillett 1956, quasi-static).

$$H = q_m r^{1/3} = \text{constant} \qquad \ldots\ldots(2)$$

(Auerbach 1891, quasi-static).

$$D/r^{2/3} = \text{constant} \qquad \ldots\ldots(3)$$

(Auerbach 1891, quasi-static ;
Andrews 1930, impact ; Tillett 1956,
both quasi-static and impact).

$$(W/A)_{\text{critical}} = \text{constant} \qquad \ldots\ldots(4)$$

(Longchambon 1934, impact).

$$(hr^{5/3})_{\text{critical}} = \text{constant} \qquad \ldots\ldots(5)$$

(Tillett 1956, impact).

Relation (2) is a reformulation of (1), with the stress q_{m} given by Hertz's theory from P, r and the elastic constants. Relation (3) would follow from (1) by Hertz's theory, if the crack were located at the boundary of the contact area, or if its distance from this boundary were assumed proportional to the diameter of the contact area. Since this geometrical similarity might not exist, and since there is some scatter in the crack location, the validity of (3) for the average values does add information. The relations (1), (4) and (5) have not been brought together before the present paper, and therefore it has gone unnoticed that they are all equivalent, in the sense that they describe the same scale effect.

Hertz's theory of the stress field under a spherical indenter is summarized, for instance, by Timoshenko (1934) and from this source one has, with some changes in notation :

$$a=\gamma P^{1/3}r^{1/3} \qquad \ldots\ldots(6)$$

$$x=\gamma^2 P^{2/3}r^{-1/3} \qquad \ldots\ldots(7)$$

$$\sigma_{\mathrm{m}}=\frac{1-2\nu_2}{3}q_{\mathrm{m}}=\frac{1-2\nu_2}{2}\frac{P}{a^2\pi} \qquad \ldots\ldots(8)$$

$$\gamma=[\tfrac{3}{4}(\kappa_1+\kappa_2)]^{1/3} \qquad \ldots\ldots(9)$$

$$\kappa_1=\frac{1-\nu_1{}^2}{E_1} \qquad \kappa_2=\frac{1-\nu_2{}^2}{E_2}. \qquad \ldots\ldots(10)$$

Here a is the radius of the circle of contact, x is the indentation depth (distance of mutual approach from contact to the state considered) and σ_{m} is the maximum value of the tensile stress in the specimen, which occurs at the boundary of the circle of contact. P, r and q_{m} have the same meaning as before, E is Young's modulus and ν is Poisson's ratio. The indices 1 and 2 refer to indenter and specimen respectively. γ and κ are abbreviations defined by (9) and (10). It is not obvious that equations (6) to (10), derived initially for a system of static, time-independent stresses and strains, are applicable to the stress field which exists momentarily during the impact. However, Hertz has made it plausible in his 'theory of impact' (cf. Love 1944) that the stress field does obey the static equations approximately, provided the impact lasts long in comparison with the time in which elastic waves travel across the highly stressed regions near the contact area. This condition is fulfilled in the experiments under discussion.

The total strain energy of specimen and indenter is, by definition,

$$U=\int_0^x P(x)dx. \qquad \ldots\ldots(11)$$

To evaluate this, P is expressed as a function of x by means of (7). The result of the integration can be written as a function of P by using (7) again, and then becomes

$$U=\tfrac{2}{5}\gamma^2 P^{5/3}r^{-1/3}. \qquad \ldots\ldots(12)$$

By using (6) one can then relate the strain energy to the contact area and finds

$$\frac{U}{a^2\pi}=\frac{2}{5\pi}\frac{P}{r}. \qquad \ldots\ldots(13)$$

This result shows the equivalence of Longchambon's relation (4) with Auerbach's relation (1), provided U differs little from W and A is proportional to $a^2\pi$. The first condition is satisfied since the ball is observed to rebound to very nearly its initial height; the second condition is satisfied since (3), in addition to (1) and (4), is valid empirically.

The kinetic energy of the ball is related to the height of fall by

$$W = 4\pi\rho g h r^3/3, \qquad \ldots\ldots(14)$$

with ρ the density and g the acceleration of gravity. Using (14), $W = U$, (12) and (6) it may be shown that Tillett's relation (5) is equivalent to (4) and thus also to (1). One may conclude that all the independently found empirical laws come to the same thing, and one may perhaps refer to them all as to Auerbach's law, after the first discoverer.

In the work done by all the different investigators, the validity of Auerbach's law has been established for about 15 different kinds of silicate glass. This number would include the samples used by some authors who have checked on Auerbach's work and found results compatible with (1), though their figures do not follow (1) quite accurately (see the reviews of Haward 1949 and Stanworth 1950). The range of indenter radii covered in the work prior to Tillett's was from $\frac{1}{2}$ mm to 15 mm; Tillett has extended this up to $r = 12{\cdot}5$ cm and confirmed the existence of an upper limit in r to the validity of Auerbach's law, which is suggested by an argument given below but was not discovered in the earlier experiments.

§ 2. INSUFFICIENCY OF THE EXPLANATION BASED ON FLAW STATISTICS

The size effect concerning the indentation hardness of glass would seem to be connected with the brittleness of this material. Metals follow the criterion of a critical stress independent of the indenter radius, both in quasi-static and in impact tests, although the yield stress measured by impact is higher (Davies 1949).

The explanation suggested in the literature for the size effect occurring with glass is based on the idea of flaws. Glass is not homogeneous with regard to its mechanical strength, and fractures start at some weak spot in the material. Often this is pictured as a sub-microscopic crack present in the glass before the load is applied (Griffith 1920), but not much is known about the nature of the flaws, though the evidence for their existence (in some form) is strong. Incidentally, the scatter in the crack location which occurs in the indentation experiments must be attributed to the varying distribution of inhomogeneities or flaws (Gibbs and Cutler 1951). As regards the explanation of the size effect itself, on the basis of flaw statistics, it runs like this. If flaws are comparatively rare, then the smaller the part of the surface which is loaded, the smaller is the chance of this part containing a particularly weak spot. Therefore, in testing the surface with small probes the average strength might be found larger than in experiments with wide-spread load. One should, however, also expect that for small probes the variation in the observed strength will be large, with the scatter increasing with decreasing probe dimensions. This effect has not been found in any of the experiments quoted. As far as qualitative statements go, all observers agree that the critical load (or height of fall or whichever is the observed quantity) is as definite for the smallest as for the larger indenters. Quantitative data for the scatter have been obtained by Tillett and they show no increase with decreasing indenter radius.

A second reason for doubting the explanation of Auerbach's size effect by flaw statistics is that it does not readily lead to the special power laws which are true empirically. These would have to be set down to a peculiar, special flaw distribution, and it seems improbable that this should have existed accidentally in all the samples of different types of glass which have been tested.

§ 3. THE ENERGY CRITERION OF RUPTURE

In the same paper in which he put forward the idea of flaws, Griffith (1920) also suggested a dynamical principle, according to which the growth of a brittle fracture is determined by the balance between the supply and the demand for energy. Energy is supplied if strain is released by the fracture, and if external forces do work upon the body during fracture; energy is demanded for the formation of new surface. The principle of this balance, which is logically quite independent of the flaw hypothesis, can by itself explain size effects, when combined with the scaling laws of geometrically similar stress fields.

The strain energy U of the Hertzian stress field as given by (12) may be re-written, so that U is expressed in terms of a and σ_{m}. The result is

$$U=\frac{2}{5}(\gamma a)^3\left[\frac{2\pi\sigma_{\mathrm{m}}}{1-2\nu_2}\right]^2. \qquad \ldots\ldots(15)$$

U is of the form $U=\epsilon a^3\sigma_{\mathrm{m}}^2$, with ϵ independent of a and σ_{m}, as one might expect. Since the cracks caused by indenters of varying r are geometrically similar, the change of the strain energy during fracture will be a certain constant fraction of U, say ηU with η a small factor independent of ϵ, a and σ_{m}. [illegible] specimen and indenter are rigidly supported while the fracture occurs, so that no work is done upon the system, then the change in strain energy is a decrease and the amount ηU becomes available. It is simplest to consider this ideal experiment involving rigid supports. (Other assumptions could be chosen, for instance, that a constant force acts between indenter and specimen, but eventually the fracture criterion would come to the same condition.) In the energetically isolated system ηU must just balance the surface energy of the crack. Since the cracks are geometrically similar, the crack surface is a constant multiple of $a^2\pi$, the contact area, say it is $\theta a^2\pi$. The surface energy of the crack is then $V=\theta a^2\pi T$, with T the surface tension (a constant). The fracture condition is

$$\eta U=V \qquad \ldots\ldots(16)$$

or

$$\frac{U}{a^2\pi}=\frac{\theta T}{\eta} \qquad \ldots\ldots(17)$$

or

$$\sigma_{\mathrm{m}}=\left(\frac{\pi\theta T}{\eta\epsilon a}\right)^{1/2}. \qquad \ldots\ldots(18)$$

The right-hand side of (17) is a product of constants; comparing with (13) and (4) one recognizes (17) as a form of Auerbach's law. The condition (18) in terms of the stress resembles that derived by Griffith for the growth of a straight crack of length $2a$; although the constant in (18) is different from that in Griffith's formula and the length is not that of the crack but the characteristic length associated with the inhomogeneous stress field. Equation (17) is essentially the form of Auerbach's law found by Longchambon, so that this is probably the form which most nearly expresses the physical meaning.

The above argument is speculative and at best gives only the form of Auerbach's law, but not the value of the constant in it. The determination of this constant would depend on a theoretical evaluation of the factor η. Also, to make the argument rigorous, it would have to be shown that the integral balance condition (16) adequately represents the differential energy balance, throughout the growth of the fracture. Regarding this last point, there is some favourable evidence in the observed propagation speed of the ring crack. In the quasi-static experiments this is sometimes so low that the crack can be followed by eye, and so far as has been established it is low compared with the velocity of elastic waves whenever Auerbach's law holds. This shows that the system is not far from mechanical equilibrium.

§ 4. LIMITS TO AUERBACH'S LAW

Even if the energy balance combined with similarity relations is the true cause of Auerbach's law, there are also other factors which control the deformation and fracture process. It might therefore be expected that Auerbach's law has only a finite range of validity.

Those forms of the law which are bound up with Hertz's theory of the stress field cannot be valid for indenters so small that a becomes comparable with r, since Hertz's equations, which are based on small deformations, then cease to apply. For steel indenters with $r = 1$ cm used on glass, the crack diameter is approximately 1 mm. From this fact and equation (3) one finds that the minimum indenter size down to which Hertz's theory can reasonably be applied is r of the order of 0·1 mm. The minimum size of Hertzian fractures found by Tolansky and Halperin (1954) on diamond was of the order of $50\,\mu = 0{\cdot}05$ mm.

Similarity relations like those of the preceding section would exist for many types of inhomogeneous stress fields, independently of whether the deformations are small or not. This suggests, tentatively to ignore the limitations given by the condition $r \gg a$ and to extrapolate (18), which does not involve r, to very small dimensions. Adjusting the constant in (18) to the critical stresses found with large indenters (where Hertz's theory applies; see below for some numerical data), and taking $a = 1\,\mu$, one finds predicted critical stresses of the order 100 000 kg cm^{-2}, i.e. the strength of the molecular bonds. Apparently under tools fine enough to produce such extremely small dents, the fracture might cease to be brittle, since the stresses demanded by the energy balance condition of a brittle crack would suffice to destroy the cohesion of the material. It is perhaps possible that this speculation gives the explanation of the phenomenon of superhardness and micro-plasticity, discovered by Smekal and co-workers (Smekal and Klemm 1951, Smekal 1953, where earlier references are given).

An upper limit in r to the validity of Auerbach's law must exist because the tensile strength of glass has a finite value, σ^* say, when determined in experiments involving large specimens and homogeneous stress fields. Auerbach's law in the form (2) predicts a decrease of the apparent tensile strength for increasing r; this decrease must ultimately conflict with the existence of σ^*. Since the value of σ^* is governed by the flaws in the material, these determine the upper limit of the range. (The more the material is weakened by flaws, the wider the range of r within which the supply of energy is the controlling condition for the fracture.) For a numerical estimate one may use some of Auerbach's results. He used glass indenters made of the same material as the specimen. For one of his glasses ('Glas II') he found $(P/r)_{\text{critical}} = 49\,\text{kg}\,\text{cm}^{-1}$ and he took $E = 675\,100\,\text{kg}\,\text{cm}^{-2}$ and $\nu = 0{\cdot}225$. (This last figure is perhaps a little low, but the error, if any, does not matter for the results below.) Using these figures and (6) one finds for $r = 1$ cm a value $a = 0{\cdot}47$ mm at the critical load and by (8) the hardness constant is $H = 10\,500\,\text{kg}\,\text{cm}^{-5/3}$. From (8) and (2) the critical tensile stress is then

$$\sigma_{\text{m}} = 1930\,r^{-1/3}, \qquad \ldots\ldots(19)$$

with r in cm and σ_{m} in kg cm^{-2}. Auerbach quotes $\sigma^* = 880\,\text{kg}\,\text{cm}^{-2}$ as the tensile strength of 'Glas II' as determined by engineering procedure. Equating this figure with σ_{m} in (19) and solving for r one finds $r = 10{\cdot}5$ cm. For r greater than 10·5 cm it cannot be the energy condition which controls the onset of fracture, but the condition of a constant minimum stress σ^* required to start a crack from a flaw. For r greater than 10·5 cm, the critical load and the crack diameter will therefore follow the laws $P \propto r^2$ and $D \propto r$. With regard to the energy, the system must now be unstable when fracture occurs, and the crack will develop catastrophically, running very fast once it gets started. The expected changes in the laws for P and D and the increase in the crack speed have all been confirmed by Tillett; with her glass, and with steel indenters, the upper limit of the range of Auerbach's law appears to lie a little lower than $r \simeq 10$ cm.

ACKNOWLEDGMENTS

The author is much indebted to Dr. H. Kolsky, Mr. D. G. Christie and Miss J. P. A. Tillett (Mrs. T. Bull) for advice and criticism, for discussion and for information on experimental data.

REFERENCES

ANDREWS, J. P., 1930, *Proc. Phys. Soc.*, **43,** 18.

AUERBACH, F., 1891, *Ann. Phys. Chem.*, **43,** 61.

DAVIES, R. M., 1949, *Proc. Roy. Soc.* A, **197,** 416.

GIBBS, P., and CUTLER, J. B., 1951, *J. Amer. Ceram. Soc.*, **34,** 200.

GRIFFITH, A. A., 1920, *Phil. Trans. Roy. Soc.*, **221,** 163.

HAWARD, R. N., 1949, *The Strength of Plastics and Glass* (London : Cleaver-Hume Press).

HERTZ, H., 1882, *Verhandlungen des Vereins zur Beförderung des Gewerbefleisses*, **61,** 449, reprinted 1895 in *Collected Works*, **1,** 174 (Leipzig : Barth).

LONGCHAMBON, L., 1934, *C. R. Acad. Sci., Paris*, **199,** 1381.

LOVE, A. E. H., 1944, *Mathematical Theory of Elasticity* (New York : Dover).

SMEKAL, A., 1953, *Acta Phys. Austriaca*, **7,** 324.

SMEKAL, A., and KLEMM, W., 1951, *Monatshefte f. Chemie (Wien)*, **82,** 411.

STANWORTH, J. E., 1950, *Physical Properties of Glass* (Oxford : Clarendon Press).

TILLETT, J. P. A., 1956, *Proc. Phys. Soc.* B, **69,** 47.

TIMOSHENKO, S., 1934, *Theory of Elasticity* (London and New York : McGraw-Hill).

TOLANSKY, S., and HALPERIN, A., 1954, *Proc. Phys. Soc.* B, **67,** 473.

TOLANSKY, S., and HOWES, V. R., 1954, *Proc. Phys. Soc.* B, **67,** 467.

Space Charge Phenomena in Liquid Dielectrics

By D. W. GOODWIN†
Department of Physics, The University of Birmingham

Communicated by K. A. Macfadyen; MS. received 19th August 1955

Abstract. By means of the electro-optical Kerr effect, distortion by space charge of the uniform potential gradient between electrodes in purified chlorbenzene has been observed. Evidence is presented suggesting that the conduction process occurs in filaments. The modified distribution of field strength is in qualitative agreement with a previous theory.

---

§ 1. INTRODUCTION

A THEORY of electrical breakdown in liquid dielectrics (Goodwin and Macfadyen 1953) suggests that field distortion is produced in a liquid at high field strengths as a result of electron emission from the cathode and the movement of ions between the electrodes. One important prediction of this theory is that the cathode field strength becomes enhanced in comparison with the anode field strength as the applied field strength is increased to that value required to produce breakdown. It is the purpose of this paper to describe investigations into such an increase in cathode field strength and to compare its magnitude with the predicted value.

Space charge effects in liquids have been observed previously (Dantscher 1931, Vafiadis 1939) using the Kerr electro-optical effect under conditions of continuously applied fields. In the experiments to be described space-charge effects, resulting from the application of rectangular voltage pulses, have been revealed using the Kerr effect. The use of pulsed voltages has been shown previously to overcome experimental difficulties associated with liquids under high electric stress.

§ 2. APPARATUS AND EXPERIMENTAL METHOD

2.1. *The Experimental Method*

The application of a strong electric field to an insulating liquid will make it doubly refracting. If now light, plane polarized in a direction at 45° to that of the electric field, passes through the liquid it will be elliptically polarized. The phase difference Δ between the two components is given by Kerr's law

$$\Delta = BlE^2 \qquad \ldots\ldots(1)$$

where l is the distance the light travels through a field of strength E; B is known as Kerr's constant. Transmission can then occur through an analyser crossed with the polarizer under zero field strength conditions. Differences in transmission across the gap can be correlated with space charge phenomena (Dillon 1930). The phase difference Δ can be measured with a Rayleigh compensator (Rayleigh 1902), a strip of photo-elastic material under cantilever strain placed before the analyser (figure 1). The phase difference in the compensator due to mechanical strain is proportional to the distance from the central axis and balances that produced in the liquid. The extinction fringe, normally coincident with the neutral axis, moves accordingly. The fringe displacement at any point between the electrodes gives a measure of the field strength at that point (equation (1)).

---

† Now at the Radar Research Establishment, Great Malvern, Worcs.

The experimental technique involves the establishment of a momentary electric field between the electrodes immersed in the liquid under observation. During this time the field distortion is observed by recording the fringe shape. The Kerr cell is illuminated with a flash tube, the period of illumination being short compared with the voltage pulse duration.

2.2. *Apparatus and Experimental Method*

A block diagram of the apparatus is given in figure 1. On closure of a mechanical switch a rectangular voltage pulse was produced by the artificial transmission line generator, with a voltage controllable up to 20 kv and a pulse duration controllable up to one millisecond, and applied to two plane parallel electrodes immersed in liquid. A delay which could be varied up to one millisecond was included in the flash tube circuit, so that the tube could be triggered at any time during the application of the voltage pulse. The flash duration was approximately 50 microseconds. Both the voltage pulse and the light flash could be monitored by means of a cathode-ray oscillograph. The optical components were mounted on an optical bench fitted with transverse slides. Accurate collimation of the incident light was essential to ensure that each part of the light beam passed through a region of uniform field strength. It was essential that the electrode faces were parallel to the direction of the light beam. Failure to be so would result in a misrepresentation of any space charge phenomena. Plane parallel electrodes were used whose edges were shaped to the Rogowski profile (Rogowski 1923) to prevent non-uniform fields over the electrode length. Auto-collimation of light reflected from facets accurately machined at right angles to the electrode faces ensured that the electrode faces were parallel to the direction of incident light.

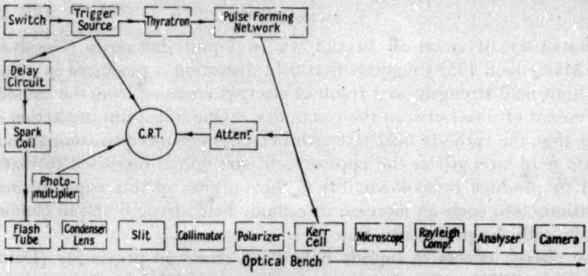


Figure 1. Block diagram of apparatus.

The two highly polished electrodes were supported in a glass cell fitted with fused-on optical windows. The electrode separation was controlled externally and measured with a microscope having a calibrated eyepiece.

The procedure for purification of the liquid adopted was similar to that used for conduction measurements (Goodwin and Macfadyen 1953).

An image of the electrodes was produced coincidental with the Rayleigh compensator plate and illumination of the compensator by the flash tube recorded photographically on Ilford 5G91 film. In normal operation the separation between two adjacent extinction fringes was greater than the width of the compensator plate. Doubly refracting plates of known phase difference were used therefore to calibrate the compensator. These plates were calibrated then by using a photo-elastic resin as the compensator plate for which the separation of adjacent fringes was less than the plate width. This calibration was carried out in monochromatic light.

§ 3. Experimental Results

3.1. *Preliminary Measurements*

Measurements were made of the fringe displacement with applied field strength for a number of non-polar liquids. In all these measurements a fixed pulse duration of 0·7 millisecond was used. Phosphor bronze electrodes of 1 cm diameter were used with a separation of 0·015 cm. The light beam was restricted to a depth of 0·1 cm, so ensuring that the optical path length in the electric field (equation 1) was equal to the diameter of the electrodes. The fringe displacement was found to be proportional to the square of the applied field strength for all the liquids used up to their respective breakdown strengths. This exhibits the validity of Kerr's law (equation (1)). Values of Kerr's constant (in e.s.u.) for these liquids are as follows: *n*-hexane, $5{\cdot}4 \pm 0{\cdot}1 \times 10^{-9}$; benzene, $5{\cdot}90 \pm 0{\cdot}05 \times 10^{-8}$; xylene, $7{\cdot}8 \pm 0{\cdot}05 \times 10^{-8}$; toluene, $7{\cdot}9 \pm 0{\cdot}05 \times 10^{-8}$.

No space charge effects were observed for applied field strengths approaching the breakdown strength of the liquid. These measurements are therefore important in that Kerr's law has been verified for them at field strengths up to 1·0 MV cm^{-1} under conditions in which no space charge existed in the liquid.

Previous calculations based on conduction and breakdown measurements (Goodwin and Macfadyen 1953) suggest that space charge phenomena should have been observed for field strengths of 1·2 MV cm^{-1}.

In the next section a possible explanation of this lack of field distortion is offered which is later supported experimentally.

3.2. *Filamentary Processes in Liquids*

When electrical breakdown occurs between the electrodes under the experimental conditions described above, their surfaces are marked by the arc discharge which follows the initial breakdown process. Only one such ' pock ' mark is observed on each electrode for each application of a voltage pulse capable of producing breakdown. This was always found to be less than 0·05 cm diameter and usually occurred on a part of the electrode surface away from the light beam. Observations made by Washburn (1933) suggest that spark channel widths in liquids are approximately 0·02 cm diameter. It is therefore likely that the breakdown process occurs within a filament of maximum diameter 0·05 cm which need not be near the electrode centre.

Further evidence for a filamentary process is offered by a consideration of conduction in liquids. The variation of the conductivity of a liquid with electrode spacing and applied field strength has been given previously (Goodwin and Macfadyen 1953).

Simple theory of ionization by collision in a liquid gives

$$i=i_c \exp(\alpha l) \qquad \ldots\ldots(2)$$

where α is the ionization coefficient of the charge carriers and i_c the density of electron current leaving the cathode. A logarithmic plot of current against electrode separation should be a family of straight lines for different field strengths. Their curvature has been attributed to recombination of carriers (Nikuradse 1930). A more detailed calculation is given in terms of the known recombination coefficient.

The general conduction equations are

$$\left.\begin{aligned} \frac{\partial N}{\partial t} &= \alpha v_n N - \frac{\partial n}{\partial x} - \beta NP \\ \frac{\partial P}{\partial t} &= \alpha v_n N + \frac{\partial p}{\partial x} - \beta NP \end{aligned}\right\} \qquad \ldots\ldots(3)$$

where N, P, n and p are the concentrations and fluxes of negative and positive charge carriers respectively; v_n and v_p are their velocities and β the recombination coefficient. Although collision ionization processes probably involve electrons, it will be assumed that values of α and β are those involving negative ions (Goodwin and Macfadyen 1953). Under equilibrium conditions, assuming the boundary conditions

$$\left.\begin{aligned} n&=n_0 \quad \text{at} \quad x=0 \quad \text{(cathode)} \\ p&=0 \quad \text{at} \quad x=l \quad \text{(anode)} \end{aligned}\right\} \qquad \ldots\ldots(4)$$

the total current is given by

$$\frac{I^2\beta}{A\epsilon\alpha v_n v_p} - I\left(1+\frac{\beta I_c}{\alpha v_n v_p \epsilon A}\right) + I_c \exp\left[\alpha l\left(1-\frac{I\beta}{A\epsilon\alpha v_n v_p}\right)\right] = 0 \qquad \ldots\ldots(5)$$

where A is the cross-sectional area in which the current flows and ϵ the permittivity. This equation can be solved for different values of I and l to give values of A, α, and I_c for different applied field strengths. For small values of αl approximations can be made, equation (5) being finally solved by successive approximations. Measurements given previously (Goodwin and Macfadyen 1953) have been analysed assuming mobilities and a recombination coefficient for negative and positive ions (Adamczewski 1937). The effective emitting area is found to be approximately $3\times10^{-6}\,\text{cm}^2$ corresponding to a filament diameter of 0·002 cm, for an electrode separation of 0·01 cm. This diameter increases linearly with increasing electrode separation and decreases slightly with increasing field strength.

Further evidence for a filamentary process is provided by an analysis of the field emission current density into a liquid from a cathode surface. The cathode current density i_c varies with a cathode field strength E_c as given by the Fowler–Nordheim relationship

$$i_c = aE_c^2 \exp(-b/E_c). \qquad \ldots\ldots(6)$$

For n-hexane, a and b are respectively 10^{-13} and 10^{-2} of their theoretical values for a metal–vacuum interface. Even larger deviations occur for electron emission into a compressed gas (Young 1950). Grey Morgan and Harcombe (1953) have suggested that these discrepancies can be overcome by assuming that emission occurs from an oxide film of work function 0·1 ev and from an emitting area of 10^{-15} cm^2. A theory of Llewellyn Jones and Grey Morgan (1953) can also explain the discrepancies by assuming emission from a thin tarnish film across which an intense electric field is maintained by the presence of positive ions at the film–liquid interface. The effective emitting area is found to be 5×10^{-4} cm^2 assuming a cathode metal work function of 4·5 ev, a film of thickness 10^{-7} cm and work function 0·11 ev, and a cathode field strength magnification of 10 due to surface irregularities. If the conduction process in a liquid is filamentary, the area of emission should correspond to the filament area at the cathode. Calculations based on a theory of spark channels in gases (Llewellyn Jones 1950) suggest that the effective cathode emitting area is 10^{-5} cm^2.

There is therefore reasonable evidence to suggest that at field strengths near to the breakdown value, the conduction process will take place in one or more filaments whose diameters may be as large as 200 μ. It is now obvious why no field distortion was observed using electrodes of 1 cm diameter. As shown in figure 2, if the effective filament diameter is l_1, in which the space charge field is E_s and if, in a length l_2 equal to $L-l_1$, the space-charge-free field is equal to the applied field strength E_a then the phase difference Δ between the two components of the light resolved at right angles is given by

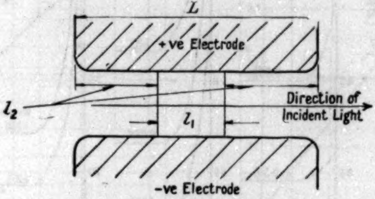


Figure 2. Filamentary conduction.

$$\Delta = B(l_1 E_s^2 + l_2 E_a^2). \qquad \ldots\ldots(7)$$

For the same phase difference to be observed with and without filamentary emission, E_s must be very much greater than E_a for the typical values of l_1 and l_2 quoted above. Breakdown would occur before any appreciable shift of the Rayleigh fringe was observed.

3.3. *Investigation of Filamentary Processes*

The most direct methods for observing the true space-charge field E_s is by the reduction of l_2 so that the phase difference Δ (equation (7)) is that due to birefringence in the filament. A search for field distortion was made therefore using electrodes whose diameters were comparable with the expected filament diameter. In order to compensate for the reduced path length (equation (1)) it was necessary to use a liquid having a larger Kerr coefficient than benzene so as to retain sufficient sensitivity. Measurements were made therefore using chlorbenzene, which has a dielectric strength comparable with that of the paraffin hydrocarbons. Highly polished and degreased brass electrodes were used. Photographic recordings of the Rayleigh fringe position as a function of field strength for two different electrode separations and for many different electrode diameters were obtained and analysed, using a travelling microscope. The fringe edges were well defined and accurate estimations of the centre of the fringe could be made. A typical set of results is shown in figure 3, for an electrode separation of 106 μ and an electrode diameter of 126 μ. The square root of the fringe displacement $y^{1/2}$ is shown as a function of the distance between the electrodes for differing field strengths. The intensification of the cathode field strength as predicted (Goodwin and Macfadyen 1953) is clearly shown. The space charge distortion observed with pulsed field was independent of the time of application of the field for values greater than 0·1 millisecond, the shortest time that could be measured due to the pulse rise time.

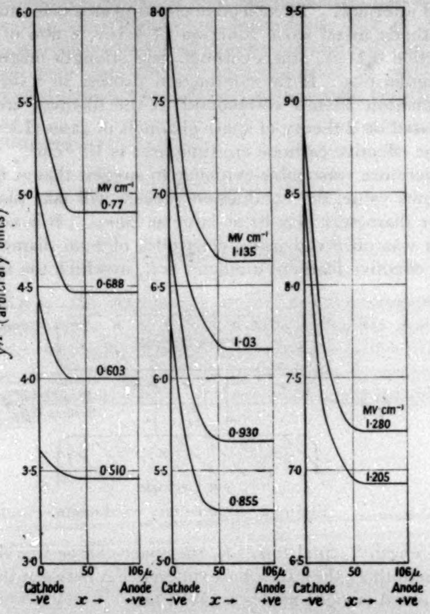


Figure 3. Fringe displacement in chlorbenzene.

As y is proportional to the phase difference Δ, a plot of $y^{1/2}$ against E (equation (1)) should be linear if Kerr's law holds. In figure 4, the average value of $y^{1/2}$, obtained by planimetric analysis is plotted as a function of the applied field strength. Deviations from Kerr's law are evident at high field strengths. It is necessary to extrapolate this curve to higher field strengths in order to convert the higher values of $y^{1/2}$ into field strengths. This is done in the following manner:

Langevin (1910) and Raman and Krishnan (1927) have shown that the Kerr constant of a liquid should be proportional to $(\epsilon-1)/(\epsilon+2)$ where ϵ is the static dielectric constant. It has also been shown (Böttcher 1952) that the dielectric constant $\Delta\epsilon$ is given by

$$\Delta\epsilon = -\frac{4\pi n\mu^4}{45(\boldsymbol{k}T)^3}\,\frac{(\epsilon_0+2)^4}{81}\,E^2 \qquad \ldots\ldots(8)$$

where n is the number of molecules per cm^3, μ their dipole moment, $\boldsymbol{k}$ Boltzmann's constant, T the absolute temperature, ϵ_0 the static dielectric constant for zero applied field strength. Kerr's constant B will be related to the value at weak fields B_0 by the relationship

$$\frac{B}{B_0} = \frac{(\epsilon-1)(\epsilon+2)}{(\epsilon_0-1)(\epsilon_0+2)} \qquad \ldots\ldots(9)$$

where $\epsilon(=\epsilon_0+\Delta\epsilon)$ is given by equation (8).

An estimate of B_0 is made from figure 4 and from this ϵ can be calculated from equation (9) for various field strengths. As shown in figure 5, $\Delta\epsilon$ is proportional to E^2 in agreement with equation (8). The dipole moment of chlorbenzene as calculated from this equation is 2×10^{-18} e.s.u. in agreement with other determinations. Equations (8) and (9) enable B to be calculated for applied field strengths greater than those used and values are shown in figure 4. All values of $y^{1/2}$ can now be converted into field strength values. The average value of $y^{1/2}$ is no longer proportional to the applied strength due to the failure of Kerr's law at high field strengths. The true field strengths are therefore averaged to check with the applied field strengths. If any deviations occur, the whole analysis is repeated and successive approximations made until agreement is obtained.

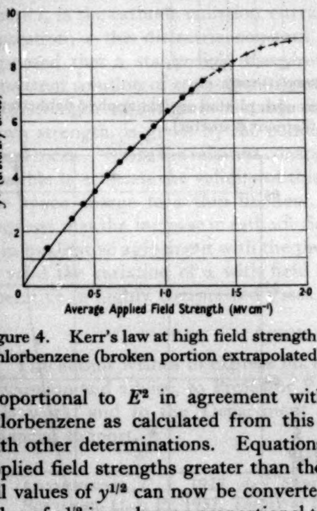


Figure 4. Kerr's law at high field strength in chlorbenzene (broken portion extrapolated).

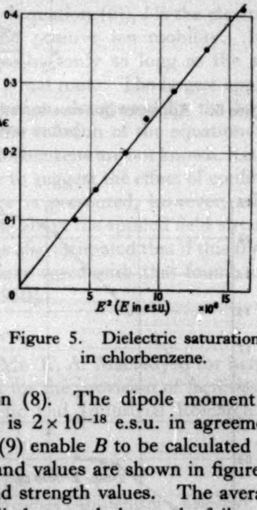


Figure 5. Dielectric saturation in chlorbenzene.

This analysis has been carried out for several electrode diameters and for two different electrode separations. Figure 6 shows the apparent increase in cathode field strength ΔE_c as a function of applied field strength E for three electrode diameters and an electrode separation of 106 μ. Other curves have been omitted for clarity. The apparent cathode field strength is related to the phase difference Δ (equation (7)) by Kerr's law. This can be expressed in the form

$$LE_c^2 = l_1E_{sc}^2 + l_2E_a \qquad \ldots\ldots(10)$$

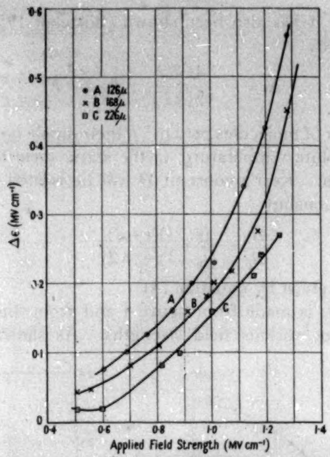


Figure 6. Apparent increase in cathode field strength plotted against applied field strength (before correction for filament diameter).

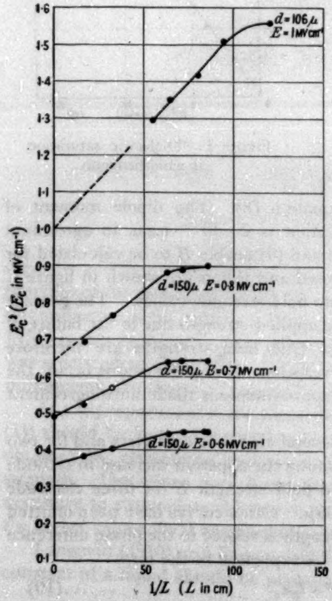


Figure 7. Variation of E_c^2 with $1/L$ for chlorbenzene.

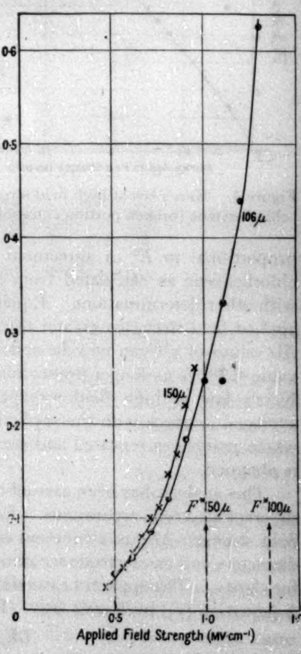


Figure 8. True increase of cathode field strength against applied field strength.

where $L = l_1 + l_2$ and E_{sc} is the true cathode field strength modified by space charge. A plot of $E_c^2 - E_a^2$ against $1/L$ gives a linear relationship until L becomes equal to l_1 (figure 7). For L equal to l_1, E_c becomes the true cathode field strength. The filament diameter varies from 120 μ for an electrode spacing of 106 μ to 180 μ for a spacing of 150 μ. Equation (10) enables the cathode field strength to be calculated and its variation with applied field strength and electrode spacing is shown in figure 8. The breakdown strengths of chlorbenzene for electrode separations of 150 and 100 μ respectively are indicated as F^* 150 μ and F^* 100 μ.

§ 4. Discussion of Results

A previous theory (Goodwin and Macfadyen 1953) showed that for large values of α, the ionization coefficient, the change in cathode field strength with applied field strength E is given by

$$\Delta E_c = \frac{2\pi i_c l \exp(\alpha l)}{\epsilon k_p E} \qquad \ldots\ldots(11)$$

where i_c is the cathode emission current density (equation (6)), l is the electrode separation, ϵ the dielectric constant and k_p the positive ion mobility. It is supposed that a stable field distribution is possible only so long as the self-consistent solution of equations (6) and (11) has a real root. The largest applied field strength which gives a stable field distribution, designated by the breakdown strength, is given by the condition that the solution of the equations has equal roots. Since the relevant constants for chlorbenzene are not known, it is not possible to estimate the validity of this theory or to suggest the effect of confining the space charge to a thin filament. Evidence is presented, however, which suggests that the increase in cathode field strength above the applied field strength is in qualitative agreement with the theory. It is also estimated that if this theory is valid the variation of α with field strength may agree with that found to be operative in highly compressed gases (Young 1950).

Acknowledgments

The author wishes to express his thanks to Mr. K. A. Macfadyen for helpful discussion and advice, to Professor P. B. Moon for the provision of facilities for this work, and to the Department of Scientific and Industrial Research for financial support.

References

Adamczewski, C. J., 1937, *Ann. Phys., Paris*, **8**, 309.
Böttcher, C. J. F., 1952, *Theory of Electric Polarisation* (Amsterdam : Elsevier).
Dantscher, J., 1931, *Ann. Phys., Lpz.*, **9**, 179.
Dillon, G. J., 1930, *Z. Phys.*, **61**, 386.
Goodwin, D. W., and Macfadyen, K. A., 1953, *Proc. Phys. Soc.* B, **66**, 85.
Grey Morgan, C., and Harcombe, D., 1953, *Proc. Phys. Soc.* B, **66**, 665.
Langevin, P., 1910, *Radium, Paris*, **7**, 249.
Llewellyn Jones, F., 1950, *Brit. J. Appl. Phys.*, **1**, 60.
Llewellyn Jones, F., and Grey Morgan, C., 1953, *Proc. Roy. Soc.* A, **218**, 88.
Nikuradse, A., 1930, *Arch. Electrotech.*, **23**, 279.
Raman, C. V., and Krishnan, K. S., 1927, *Proc. Roy. Soc.* A, **117**, 1.
Rayleigh, Lord, 1902, *Phil. Mag.*, (6), **4**, 678.
Rogowski, W., 1923, *Arch. Electrotech.*, **12**, 1.
Vafiadis, G., 1939, *Ann. Phys., Lpz.*, **35**, 23.
Washburn, H. W., 1933, *Physics*, **4**, 29.
Young, D. R., 1950, *J. Appl. Phys.*, **21**, 222.

The Measurement of the Energy Gap of Semiconductors from their Diffuse Reflection Spectra

BY P. D. FOCHS
Research Laboratory, Associated Electrical Industries, Ltd., Aldermaston, Berks.

Communicated by E. Billig; MS. received 25th August 1955

Abstract. The diffuse reflection spectrum of a powdered semiconductor is characterized by an increase in the diffuse reflectivity as the individual particles become relatively transparent. This increase has a definite linear region of greatest slope which is attributed to an exponential drop in the absorption coefficient. The onset of this exponential drop is suggested as a more universal method of determining absorption edges, and when applied to the interpretation of diffuse reflection spectra gave the following energy gaps (at room temperature): amorphous selenium 1·86 ev, metallic selenium 1·74 ev, silicon 1·20 ev, and germanium 0·69 ev.

---

§ 1. INTRODUCTION

THE simple picture of the distribution of energy states in a semiconductor on the two sides of the forbidden energy gap suggests the existence of a well-defined edge to the absorption curve. In reality, the edge is masked by a tail extending into the energy gap due to the influence of lattice vibrations (e.g. Cheeseman 1952), and consequently most estimates of its position are usually made on a somewhat arbitrary basis. A more logical method of determining the edge is to fit the adjacent short wavelength region of the edge to a theoretical absorption curve based on a particular distribution of energy states and then extrapolate to zero absorption (Seiwert 1949, Smith 1954, cf. Macfarlane and Roberts 1955).

An alternative approach is to study the tail of the absorption curve, since this might be expected to have a simple exponential drop with decreasing photon energy E. Experimental evidence is provided by Urbach (1953), who found that the slope of the absorption curve, $d(\log \alpha)/dE$ where α is the absorption coefficient, was equal to $1/\boldsymbol{k}T$ for AgBr over a range of temperatures, and was within a factor of two of $1/\boldsymbol{k}T$ for AgCl, Ge, TiO_2 and CdS; similar slopes are found in absorption curves for germanium and silicon (Dash, Newman and Taft 1955). Thus, it seems very probable that most semiconductors have this characteristic exponential drop in absorption coefficient, provided that it is not masked by impurity or free-carrier absorption. The onset of this drop is therefore suggested as a more universal method of deducing an absorption edge from such curves without being dependent on the actual shape of the main absorption band. Accurate measurement of the tail by the transmission method is, however, difficult, due in part to the requirement of comparatively large perfect crystals, up to 1 cm thick.

The purpose of this paper is to show that the measurement of the diffuse reflection spectrum of the powdered material is particularly suited to the determination of absorption edges, since it enables much of the general character of the tail to the absorption curve to be brought out in a single run with much less material ($\sim 2\times 10^{-2}\,\text{cm}^3$) and the minimum of specimen preparation. The original crystals can therefore be much smaller which is a distinct advantage in the preliminary study of new materials.

§ 2. EXPERIMENTAL

The crystalline material was ground in an agate pestle and mortar, and then put through one or more sieves to restrict somewhat the size distribution of the particles in a given sample. The powder filled a simple container made from a 2 BA washer and a glass cover slip, though the specimen thickness (1·2 mm) is not in fact important provided the transmission coefficient is negligible. The powder was placed at the focus of a normally incident convergent beam of monochromatic radiation. Approximately 25% of the diffusely reflected radiation was collected by an aluminized concave mirror of N.A. 0·5, and refocused on to the detector; the powder and detector were placed as close as possible to the centre of curvature of the mirror in order to minimize the various aberrations. Vacuum photocells and a balanced electrometer unit were used as the detector for the region 0·22 to 1·1 μ, while a PbS cell and homodyne amplifier enabled measurements to be extended to about 3 μ. A Hilger large-aperture monochromator was the basic dispersing system throughout this range, while to minimize the scattered radiation in the ultra-violet a double monochromator was set up by the addition of a Hilger–Barfit monochromator. Although comparatively wide slit widths were used (0·2 mm), they had no significant effect on the actual spectra.

The reflectivity of the specimen was measured relative to a suitable 'standard', which clearly should not superimpose its own characteristic spectrum on that of the specimen. Thus, CaF_2 or LiF was used for the ultra-violet region, MgO for the visible, and silver dust or the specular reflection from an aluminium mirror for the near infra-red. (The silver dust was inferior beyond about 2·5 μ but this may have been due to the very small particle size.) No attempt was made to measure absolute reflectivities, since it was found that the reflectivity of both the specimen and the 'standard' may easily vary by several per cent, due to ageing of the powders, water content, packing factor, etc., and the resultant correction factor does not in general materially affect the shape of the spectra.

§ 3. RESULTS

Typical reflection spectra for five semiconductors are plotted on a reciprocal wavelength scale in figure 1. Each curve possesses a region of relatively constant reflectivity at the shorter wavelengths due to the surface reflectivity of the material, since, provided the particles are completely absorbing, the absorption coefficient has little effect until the centre of the main absorption band is approached.

At longer wavelengths, the decreasing absorption coefficient results in the individual particles becoming relatively transparent, leading to multiple reflections and refractions in the powder and a net increase in the diffuse reflectivity.

Subsequent increases are largely dependent on the value of the absorption coefficient, and when this is very low the reflectivity approaches unity.

All five curves in figure 1 have a distinctly linear region of greatest slope and the actual slopes agree within ±25% of their mean value. Considering the experimental conditions, this suggests a very close correlation with the observation that the slope of the absorption curve in the energy gap is probably the same in such materials. Thus, the onset of the linear increase in reflectivity may be correlated with the onset of the tail of the absorption curve and, from the suggestion in the Introduction, the onset of the linear increase may therefore be used to obtain an absorption edge from such spectra—provided, of course, that it can be clearly resolved from the other characteristic regions of the spectra.

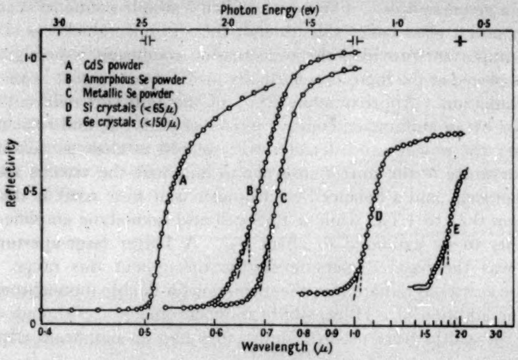


Figure 1. Diffuse reflection spectra.

Applying this 'definition' to figure 1 the following absorption edges can be deduced: amorphous selenium at about 6650 Å (≡1·86 ev), metallic selenium at about 7100 Å (≡1·74 ev), silicon at about 1·03 μ (≡1·20 ev), and germanium at about 1·80 μ (≡0·69 ev). A discussion on the existing data will be given later.

The effect of using smaller particle sizes in silicon is to bring out detail in the region of higher absorption coefficients at the sacrifice of detail in the lower absorption coefficient region, since the curves are in fact dependent on some function of the product of absorption coefficient and particle size (Johnson 1952). Thus, in figure 2, curve A, the linear increase in reflectivity tends to be absorbed in the upper portion of the graph (cf. figure 2, curves B and C). With larger particles, the surface reflectivity region extends to longer wavelengths and tends to mask the absorption edge (curve C); also, the initial increase in reflectivity, due to multiple reflections, becomes more pronounced. (The higher surface reflectivity in curve C arises because the actual number of particles irradiated is considerably reduced and thus the scattered radiation is no longer sufficiently diffuse.)

The spectrum of germanium (figure 1, curve E) is characterized by two pronounced increases in slope at about 1·53 and 1·74 μ, followed by an absorption edge at about 1·80 μ (≡0·69 ev). The first increase in slope bears a strong resemblance to the increase obtained with the larger silicon particles in figure 2, curve C. In germanium, however, the use of smaller particles has only served to emphasize the presence of this abrupt increase at about $1{\cdot}53\,\mu$, and it must therefore be regarded as evidence of a distinct 'knee' in the actual absorption curve. Independent evidence of such a knee can be seen in the actual measurements of Briggs (1952) and in the recent absorption curve of Dash, Newman and Taft (1955).

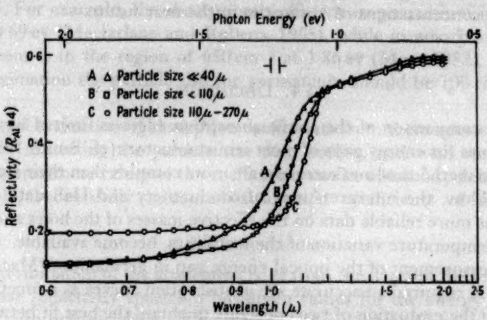


Figure 2. Effect of particle size in silicon.

In the case of CdS (figure 1, curve A) the material used was in the form of precipitated powder, and it was not possible to determine an unambiguous absorption edge with the above definition. An alternative method of deducing an edge from the intersection of the two linear portions of figure 1, curve A, has been used for ZnS (Bube 1953), but it is clearly limited in its applicability and is difficult to justify theoretically.

Added impurities do not significantly affect the characteristic spectra of the host material, although clearly as the proportion of impurity is increased, so does its own reflection spectrum become more pronounced. On the other hand, if the impurity is taken up in the lattice, then the reflection spectra are sensitive to much lower impurity contents. This is illustrated by three curves (figure 3)

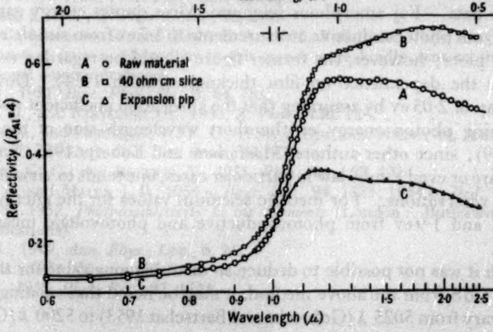


Figure 3. Effect of impurities in silicon (particle size $<65\,\mu$).

obtained in the 'history' of a silicon single crystal grown by the Czochralski (1917) technique: curve A represents the raw material with known impurities totalling about 0·002%, B a slice of resistivity 40 ohm cm p-type, and C the expansion pip. As expected, B has the lowest absorption in the infra-red, while C has the highest due largely to segregation of various impurities in the raw material to the last part of the ingot to form. It is hoped to correlate the diffuse reflectivity with known concentrations of impurities in the near future.

§ 4. DISCUSSION

A direct comparison of the above absorption edges is limited by the lack of accurate values for energy gaps of most semiconductors (cf. Smith 1954) because the two main methods of measurement are more complex than their simple theory suggests. Thus, the interpretations of conductivity and Hall data are subject to revision as more reliable data on the effective masses of the holes and electrons, and on the temperature variation of the mobilities, become available. Similarly, the latest measurement of the optical energy gap in germanium (Macfarlane and Roberts 1955) required an accurate set of absorption curves as a function of temperature and the evaluation of two constants to obtain the best fit between theory and experiment; (although the general agreement is good, the theory does not adequately represent the measurements on the long wavelength side of the absorption edge or at low temperatures where the complex nature of the absorption curve is apparent).

The recent analysis of conductivity and Hall data by Morin and Maita (1954 a) for germanium gives a value of 0·68 ev for the energy gap at 300°K, while Macfarlane and Roberts' data give 0·655 ev. Thus, the value found from the above measurements (0·69 ev) can be considered to be in reasonable agreement with the existing data.

Morin and Maita (1954 b), in a similar analysis of the conductivity and Hall measurements on silicon, obtain an energy gap of 1·10 ev (300°K). An average figure for the optical energy gap is also about 1·1 ev (Moss 1952). Thus, in this case, the value obtained above is about 0·1 ev higher than the existing data.

In both varieties of selenium, the above results are well within the spread of the available data. For amorphous selenium, Moss quotes energy gaps varying from 2·5 ev from photoconductive measurements to 1·7 ev from simple resistivity–temperature plots; however, the former figure should be regarded as an upper limit due to the dependence on film thickness (Smith 1954). Dowd (1951) deduced a gap of 2·05 ev by assuming that the absorption coefficient rises linearly with increasing photon energy on the short wavelength side of the edge (cf. Seiwert 1949); since other authors (Macfarlane and Roberts 1955, Smith 1954) prefer a square or even a cube law in particular cases, one tends to view the method with certain reservations. For metallic selenium values for the energy gap vary between 1·5 and 1·9 ev from photoconductive and photovoltaic measurements respectively.

Although it was not possible to deduce an unambiguous value for the absorption edge in CdS from the above method, it may be noted that existing estimates of the edge vary from 5025 Å (Gobrecht and Bartschat 1953) to 5200 Å (Gottesman and Ferguson 1954).

Measurements on a number of different silicon and germanium specimens have shown that the diffuse reflection method gives reproducible results within ± 1% and that it is not critically dependent on particle size. It was, however, found necessary to use larger particle sizes for the materials of lower energy gap in order to bring out the actual absorption edge most clearly, because the absorption edges did not correspond to a constant absorption coefficient in the various materials. For example, the absorption coefficient in germanium is only about 6 cm^{-1} at 0·69 ev (Macfarlane and Roberts 1955), while in amorphous selenium the coefficient is in the region of 650 cm^{-1} at 1·86 ev (Moss 1952). Thus to a first approximation the particle size for germanium should be 100 times that for selenium.

The actual evaluation of absorption coefficients from these spectra is difficult, although by restricting the particle size distribution, and with various simplifications, approximate values of the coefficients can be obtained; Johnson (1952), for example, successfully measured absorption coefficients of the order of 10^{-1} cm^{-1}.

§ 5. CONCLUSIONS

The determination of the absorption edge from the onset of the linear increase in the diffuse reflectivity spectrum has given values for the energy gap in good agreement with existing data. The measurement of the reflection spectrum of the powdered material does not in principle yield any more information than could be obtained from the accurate measurement of the actual absorption curve. In practice, however, it enables one to explore the regions of very low absorption coefficients more readily with much less material.

ACKNOWLEDGMENTS

The author is indebted to J. B. Willis and D. B. Gasson for preparing the germanium and silicon crystals respectively, to Dr. E. Billig for his advice, and to Dr. T. E. Allibone and the Admiralty for permission to publish this paper.

REFERENCES

BRIGGS, H. B., 1952, *J. Opt. Soc. Amer.*, **42**, 686.

BUBE, R. H., 1953, *Phys. Rev.*, **90**, 70.

CHEESEMAN, I. C., 1952, *Proc. Phys. Soc.* A, **65**, 25.

CZOCHRALSKI, J., 1917, *Z. Phys. Chem.*, **92**, 219.

DASH, W. C., NEWMAN, R., and TAFT, E. A., 1955, *Bull. Amer. Phys. Soc.*, **30**, 53. (The absorption curves referred to in this paper were kindly sent to the author in a personal communication.)

DOWD, J. J., 1951, *Proc. Phys. Soc.* B, **64**, 783.

GOBRECHT, H., and BARTSCHAT, A., 1953, *Z. Phys.*, **136**, 224.

GOTTESMAN, J., and FERGUSON, W. F. C., 1954, *J. Opt. Soc. Amer.*, **44**, 368.

JOHNSON, P. D., 1952, *J. Opt. Soc. Amer.*, **42**, 978.

MACFARLANE, G. G., and ROBERTS, V., 1955, *Phys. Rev.*, **97**, 1714.

MORIN, F. J., and MAITA, J. P., 1954 a, *Phys. Rev.*, **94**, 1525; 1954 b, *Ibid.*, **96**, 28.

MOSS, T. S., 1952, *Photoconductivity in the Elements* (London: Butterworths Scientific Publications).

SEIWERT, R., 1949, *Ann. Phys., Lpz.*, **6**, 241.

SMITH, R. A., 1954, *Physica*, **20**, 910.

URBACH, F., 1953, *Phys. Rev.*, **92**, 1324.

Impurity Scattering in Semiconductors

By R. MANSFIELD
Bedford College, London

MS. received 11*th August* 1955

Abstract. The theory of impurity scattering in semiconductors has been developed, and the combination of impurity and lattice scattering has been considered for the general case of any degree of degeneracy of the charge carriers. A comparison is made between theory and experimental results obtained with germanium and indium antimonide.

---

§ 1. Introduction

THE theory of impurity scattering in non-ionic semiconductors has been discussed by Debye and Conwell (1954) and a formula has been given, which was derived by Herring, for the case of scattering by impurity centres having a screened Coulomb field, when the charge carriers are not degenerate. Mott (1936) has considered the scattering of electrons by a screened Coulomb field in the case of metals when the electron gas is degenerate. Similarly, the combination of impurity and lattice scattering has been considered for the limiting cases of degeneracy. Frequently, however, impurity scattering has to be considered when it cannot be assumed that the charge carriers are either degenerate or non-degenerate. The object of this paper is to consider the general case of arbitrary degeneracy and the combination of lattice and impurity scattering under these conditions.

§ 2. Theory

2.1. *Impurity Scattering*

The expression for the conductivity σ is as follows:

$$\sigma=\frac{16\sqrt{2}\mathbf{e}^2\pi m^{*1/2}}{3\mathbf{h}^3}(\mathbf{k}T)^{3/2}\int_0^\infty \tau\eta^{3/2}\frac{\exp(\eta-\eta^*)\,d\eta}{\{\exp(\eta-\eta^*)+1\}^2} \qquad \ldots\ldots(1)$$

where η and η^* are the reduced energy and reduced chemical potential of the electrons. Following the theory derived by Mott (1936), the relaxation time is given by

$$\frac{1}{\tau}=Nv\int_0^\pi(1-\cos\theta)\,I(\theta)\,2\pi\sin\theta\,d\theta.$$

N is the concentration of impurity centres. For scattering by a screened Coulomb potential given by

$$\phi=\frac{\mathbf{e}}{\epsilon r}\,\mathrm{e}^{-qr} \qquad \ldots\ldots(2)$$

where ϵ is the dielectric constant

$$I(\theta)=\left[\frac{\mathbf{e}^2/\epsilon}{m^*v^2(1-\cos\theta)+\gamma^2}\right]^2, \qquad \gamma^2=\frac{q^2\mathbf{h}^2}{8\pi^2m^*}$$

hence

$$\frac{1}{\tau}=Nv2\pi\left\{\frac{\boldsymbol{e}^2}{\epsilon m^* v^2}\right\}^2 f(x)$$

where

$$f(x)=\log(1+x)-\frac{x}{1+x}$$

$$x=\frac{16\pi^2 m^{*2} v^2}{q^2 \boldsymbol{h}^2}$$

and

$$\tau=\frac{\sqrt{(2m^*)}\,\epsilon^2(\eta \boldsymbol{k}T)^{3/2}}{\pi N \boldsymbol{e}^4 f(x)}\,.$$

Substitution of this value of τ in equation (1) gives

$$\sigma_{\mathrm{I}}=\frac{32\epsilon^2 m^*(\boldsymbol{k}T)^3}{3N\boldsymbol{e}^2\boldsymbol{h}^3}\int_0^\infty \frac{\eta^3\exp(\eta-\eta^*)\,d\eta}{\{\exp(\eta-\eta^*)+1\}^2 f(x)}\,. \qquad \ldots\ldots(3)$$

To evaluate the integral it is necessary to assume that, since $f(x)$ is a slowly varying function of η, it may be taken outside the integral and calculated from the value of η which makes the integral a maximum. Equation (3) then becomes after integrating by parts

$$\sigma_{\mathrm{I}}=\frac{32\epsilon^2 m^*(\boldsymbol{k}T)^3}{N\,\boldsymbol{e}^2\boldsymbol{h}^3 f(x)}\int_0^\infty \frac{\eta^2\,d\eta}{\exp(\eta-\eta^*)+1} \qquad \ldots\ldots(4)$$

where $\bar{\eta}$, the value of η to be used to calculate $f(x)$ is given by the equation

$$(\bar{\eta}-3)\exp(\bar{\eta}-\eta^*)=\bar{\eta}+3. \qquad \ldots\ldots(5)$$

This equation may be solved graphically.

Equation (4) is similar to the expression derived by Johnson and Lark-Horovitz (1947), who have considered scattering by a Coulomb field for any degeneracy.

In order to calculate the potential field near an ionized impurity centre and hence to determine the screening constant q in equation (2) it is necessary to solve Poisson's equation. For the conduction band, let n_0 be the concentration of electrons in the unperturbed lattice and n the concentration of electrons in the perturbed lattice, then, since n_0 is also the density of positive charge, Poisson's equation becomes

$$\nabla^2(\phi)=-\frac{4\pi\boldsymbol{e}}{\epsilon}(n_0-n).$$

If $F(\Phi)=n_0$ and $F(\Phi+\phi)=n$ assume that $F(\Phi+\phi)=F(\Phi)+\phi F'(\Phi)$ where $\boldsymbol{e}\Phi=$ maximum energy of the electrons $=\eta^*\boldsymbol{k}T$ and

$$F(\Phi)=\frac{4\pi(2m^*\boldsymbol{k}T)^{3/2}}{\boldsymbol{h}^3}\int_0^\infty\frac{\eta^{1/2}d\eta}{\exp(\eta-\eta^*)+1}=\frac{4\pi(2m^*\boldsymbol{k}T)^{3/2}}{\boldsymbol{h}^3}f_{1/2}(\eta^*)$$

also

$$F'(\Phi)=\frac{4\pi(2m^*\boldsymbol{k}T)^{3/2}}{\boldsymbol{h}^3}f_{1/2}'(\eta^*)\frac{\boldsymbol{e}}{\boldsymbol{k}T}$$

then

$$\nabla^2(\phi)=\frac{4\pi\boldsymbol{e}}{\epsilon}\phi F'(\Phi)=q^2\phi.$$

The solution of Poisson's equation then becomes

$$\phi=\frac{\boldsymbol{e}}{\epsilon r}\,\mathrm{e}^{-qr}$$

where

$$q^2=\frac{16\pi^2\boldsymbol{e}^2(2m^*)^{3/2}\sqrt{(\boldsymbol{k}T)}}{\epsilon\boldsymbol{h}^3}f_{1/2}'(\eta^*).$$

Limiting case $\eta^* \gg 0$.

When $\eta^* \gg 0$

$$f_{1/2}'(\eta^*) = \eta^{*1/2}, \qquad n_0 = \left\{\frac{2m^* \boldsymbol{k} T \eta^*}{\boldsymbol{h}^2}\right\}^{3/2} \frac{8\pi}{3}$$

and

$$q^2 = \frac{16\pi^2 m^* \mathbf{e}^2}{\epsilon^2 \boldsymbol{h}^2} \left(\frac{3n_0}{\pi}\right)^{1/3}.$$

This is the value obtained by Mott (1936) except for the introduction of the dielectric constant. If $\eta^* > 6$ and $n_0 = N$ then

$$f_2(\eta^*) = \int_0^\infty \frac{\eta^2 \, d\eta}{\exp(\eta - \eta^*) + 1} = \frac{\eta^{*3}}{3}.$$

The equation for the conductivity becomes

$$\sigma_{\mathrm{I}} = \frac{3\boldsymbol{h}^3 \epsilon^2 n_0}{16\pi^2 \mathbf{e}^2 m^{*2} f(x)} \qquad \text{......(6)}$$

with

$$x = \left(\frac{\boldsymbol{h}}{\mathbf{e}}\right)^2 \frac{\epsilon}{m^*} \left(\frac{3n_0}{8\pi}\right)^{1/3}.$$

This equation shows that for complete degeneracy, the mobility becomes dependent on the concentration of scattering centres only through the term $f(x)$. This arises because although the number of collisions between charge carriers and impurity centres is increased, the energy of the charge carriers also increases, resulting in a greater amount of small angle scattering.

Limiting case $\eta^* \ll 0$.

$$f_{1/2}'(\eta^*) = \frac{\sqrt{\pi}}{2} \exp \eta^*, \quad n_0 = \frac{2(2\pi m^* \boldsymbol{k} T)^{3/2}}{\boldsymbol{h}^3} \exp \eta^*, \quad q^2 = \frac{4\pi \mathbf{e}^2}{\epsilon \boldsymbol{k} T} n_0.$$

$$f_2(\eta^*) = 2 \exp \eta^* = \frac{n_0 \boldsymbol{h}^3}{(2\pi m^* \boldsymbol{k} T)^{3/2}}$$

and equation (5) gives $\bar{\eta} = 3$.
Hence the mobility

$$u = \frac{\sigma}{n_0 \mathbf{e}} = \frac{2^{7/2} \epsilon^2 (\boldsymbol{k} T)^{3/2}}{N \pi^{3/2} m^{*1/2} \mathbf{e}^3 f(x)}$$

where

$$x = \frac{6m^* \epsilon (\boldsymbol{k} T)^2}{\hbar^2 \pi n_0 \mathbf{e}^2}.$$

This is the expression derived by Herring (Debye and Conwell 1954).

General case.

The general equation for the conductivity due to impurity scattering can be written as

$$\sigma_{\mathrm{I}} = \frac{32 \epsilon^2 m^* (\boldsymbol{k} T)^3 f_2(\eta^*)}{N \mathbf{e}^2 \boldsymbol{h}^3 f(x)} \qquad \text{......(7)}$$

where

$$f(x) = \log(1 + x) - \frac{x}{1 + x}$$

$$x = \left(\frac{4\pi m^* v}{q \boldsymbol{h}}\right)^2 = \frac{\bar{\eta} (\boldsymbol{k} T)^{1/2} \epsilon \boldsymbol{h}}{\mathbf{e}^2 (2m^*)^{1/2} f'_{1/2}(\eta^*)}.$$

2.2. *Combination of Impurity and Lattice Scattering*

The expression for the total conductivity can be written as follows:

$$\sigma = \frac{\sigma_L}{\log\{1+\exp\eta^*\}}\int_0^\infty \frac{\eta^3\exp(\eta-\eta^*)\,d\eta}{(\eta^2+\alpha)\{\exp(\eta-\eta^*)+1\}^2} \qquad \ldots\ldots(8)$$

where σ_L is the conductivity which would arise if lattice scattering only occurred at the appropriate value of η^*, and

$$\alpha = \frac{\sigma_L}{\log(1+\exp\eta^*)}\,\frac{3\boldsymbol{h}^3Ne^2}{32\epsilon^2m^*(\boldsymbol{k}T)^3}f(x)$$

$$= \frac{3f_2(\eta^*)}{\log\{1+\exp\eta^*\}}\,\frac{\sigma_L}{\sigma_I}. \qquad \ldots\ldots(9)$$

Assuming that $f(x)$ can be treated as a constant, α was calculated from equation (9), and the integral in equation (8) determined, for different values of η^* and σ_L/σ_I. Figure 1 illustrates the correcting factor $F=(\rho_L+\rho_I)/\rho$ which has to be applied to the sum of the resistivities ρ_L due to lattice scattering and ρ_I due to impurity scattering, to obtain the total resistivity ρ. When $\eta^* \gg 0$ the total resistivity is given by the sum of the resistivities and when $\eta^* \ll 0$ the theory of Jones (1951) and Johnson and Lark-Horovitz (1951) can be used.

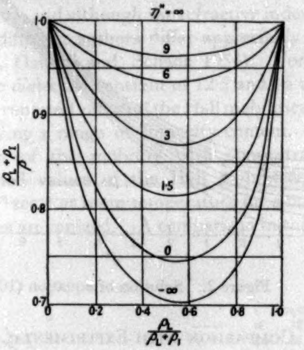


Figure 1. Combination of lattice and impurity scattering.

In order to apply these results it must be remembered that σ_I should be calculated from equation (7) using the value of $\eta=\bar{\eta}$ which makes the integrand of equation (8) a maximum. The appropriate value of $\bar{\eta}$ is given by the solution of the equation

$$\frac{\bar{\eta}^2+3\alpha}{\bar{\eta}^3+\alpha\bar{\eta}} = \tanh\left(\frac{\bar{\eta}-\eta^*}{2}\right). \qquad \ldots\ldots(10)$$

The solution of this equation for the various values of η^* and σ_L/σ_I which have been considered is given in the table and is illustrated in figure 2.

The error introduced by assuming that $f(x)$ can be treated as a constant in the integral was investigated by determining the value of the integral for a few values of η^* and σ_L/σ_I, choosing the special case of $x=10\eta$. The difference between the true values and the approximate values was less than 3%. The error, however, becomes appreciable if the value of $\bar{\eta}$ determined from equation (5) instead of equation (10) is used.

Values of $\bar{\eta}$ satisfying equation (10)

σ_L/σ_I \ η_2^*	9	6	3	1·5	0
∞	9·643	6·928	4·572	3·725	3·244
9	9·600	6·878	4·464	3·575	3·044
4	9·558	6·804	4·366	3·446	2·885
2	9·502	6·726	4·245	3·292	2·703
1	9·433	6·630	4·099	3·112	2·494
0·5	9·364	6·533	3·950	2·927	2·281
0·25	9·306	6·452	3·820	2·760	2·082
0·111	9·263	6·388	3·708	2·607	1·888
0	9·218	6·320	3·575	2·391	1·543

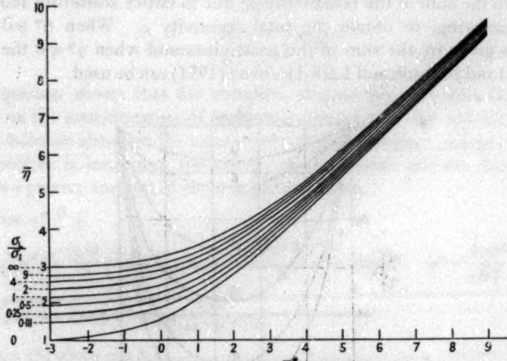


Figure 2. Solution of equation (10).

§ 3. Comparison with Experimental Results

For a comparison with experimental results it is useful to consider samples in which the impurities are all ionized over a considerable range of temperatures, then the concentration of scattering centres is the same as the concentration of charge carriers and scattering by neutral impurities can be neglected. If the residual conductivity is known for the limiting case $\eta^* \gg 0$ then it is possible to calculate the effective mass of the charge carriers using equation (6), provided the dielectric constant and concentration of charge carriers is known. The latter quantity can usually be calculated from the Hall constant and the former determined from the refractive index, if a direct measurement of the dielectric constant has not been made. Having determined a value of m^*, it is possible to calculate η^* using the expression

$$n_0 = \frac{4\pi(2m^* kT)^{3/2}}{h^3} \int_0^\infty \frac{\eta^{1/2}\, d\eta}{\exp(\eta - \eta^*) + 1}.$$

$f_{\frac{1}{2}}(\eta^*)$ and $f_{1/2}'(\eta^*)$ may then be determined from tables of these functions (Johnson and Shipley 1953, McDougall and Stoner 1938). Hence the conductivity arising from impurity scattering for any temperature may be calculated.

Over the temperature range in which impurity and lattice scattering are comparable it is necessary to calculate σ_{L} using the expression

$$\sigma_{\mathrm{L}} = \frac{2(2\pi m^* kT)^{3/2}}{h^3}\, eu \log(1+\exp\eta^*)$$

where u is the lattice mobility at temperature T when the charge carriers are not degenerate. The ratio $\sigma_{\mathrm{L}}/\sigma_{\mathrm{I}}$ may then be determined, and hence the appropriate value of $\bar{\eta}$ obtained and a new value of σ_{I} calculated. It is generally unnecessary to redetermine a value of $\bar{\eta}$ using the new value of σ_{I}. The correcting factor F can now be determined from figure 1 and the value of the total resistivity obtained.

This procedure was adopted in the case of germanium and indium antimonide. Debye and Conwell (1954) have reported measurements on an impure n-type sample of germanium (specimen 58) in which the conductivity is constant over a considerable range of temperatures. Agreement is obtained between theory and experiment if it is assumed that $\epsilon = 16{\cdot}1$ and $m^*/m = 0{\cdot}29$.

In the case of indium antimonide a direct determination of the dielectric constant has not been made, and although the refractive index has been measured, the values obtained by different authors differ appreciably (Breckenridge *et al.* 1954, Avery *et al.* 1954, Oswald and Schade 1954). For a comparison with experiment, values of the dielectric constant of 12·5 and 16 were chosen. Barrie and Edmond (1955) have reported values of the Hall mobilities at room temperature of n-type specimens having a range of impurity content. Figure 3 illustrates the theoretical variation of the mobility with concentration and gives the experimentally determined values of the Hall mobilities, assuming a lattice mobility of 80 000 $\mathrm{cm^2\,v^{-1}\,sec^{-1}}$ at room temperature for a non-degenerate sample and that all the impurities are ionized. A comparison in this case is complicated by the uncertainty of the relation between the Hall and conductivity mobilities and also in the actual value of the concentration of charge carriers. For the purer specimens the concentration of scattering centres is less than the concentration

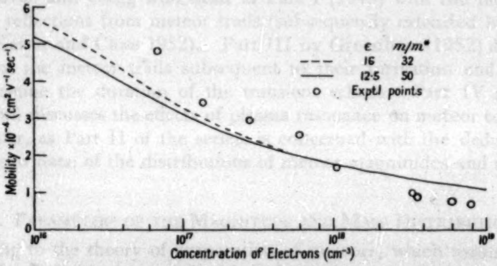


Figure 3. Variation of mobility with concentration of electrons in InSb. The experimental points are those of Barrie and Edmond (1955).

of charge carriers, since at room temperature the contribution of intrinsic electrons will become important for n less than 10^{17} cm^{-3}. In this case the calculated value of the mobility assuming $n_0 = N$ will be too small. When the values of ϵ and $\mathbf{m}/m^*$ used in figure 3 are applied to the n-type specimens studied by Madelung and Weiss (1954) reasonable agreement is obtained. It is seen that the conductivity depends mainly on the product $\epsilon \mathbf{m}/m^*$ and consequently it is not possible to decide on the correct value of $\mathbf{m}/m^*$ from these measurements.

REFERENCES

AVERY, D. G., GOODWIN, D. W., LAWSON, W. D., and MOSS, T. S., 1954, *Proc. Phys. Soc.* B, **67,** 761.

BARRIE, R., and EDMOND, J. T., 1955, *J. Electronics*, **1,** 161.

BRECKENRIDGE, R. G., BLUNT, R. F., HOSLER, W. R., FREDERIKSE, H. P. R., BECKER, J. H., and OSKINSKY, W., 1954, *Phys. Rev.*, **96,** 571.

DEBYE, P. P., and CONWELL, E. M., 1954, *Phys. Rev.*, **93,** 693.

JOHNSON, V. A., and LARK-HOROVITZ, K., 1947, *Phys. Rev.*, **71,** 374; 1951, *Ibid.*, **82,** 977.

JOHNSON, V. A., and SHIPLEY, F. M., 1953, *Phys. Rev.*, **90,** 523.

JONES, H., 1951, *Phys. Rev.*, **81,** 149.

MCDOUGALL, J., and STONER, E. C., 1938, *Phil. Trans. Roy. Soc.* A, **237,** 67.

MADELUNG, O., and WEISS, H., 1954, *Z. Naturf.*, **9a,** 527.

MOTT, N. F., 1936, *Proc. Camb. Phil. Soc.*, **32,** 281.

OSWALD, F., and SCHADE, R., 1954, *Z. Naturf.*, **9a,** 611.

Characteristics of Radio Echoes from Meteor Trails
II: The Distribution of Meteor Magnitudes and Masses

By I. C. BROWNE, K. BULLOUGH, S. EVANS AND T. R. KAISER
Jodrell Bank Experimental Station, University of Manchester

MS. received 17th August 1955

Abstract. Magnitude and mass distributions are deduced from radio echo data for sporadic meteors and for various meteor showers. The frequency of sporadic meteors is found to increase by 2·5 times per unit increase in magnitude over the range 0^{m}–10^{m}. This corresponds to an inverse square law for the differential mass distribution. Results obtained for shower meteors in the magnitude ranges 0^{m} to 2^{m} and 5^{m} to 7^{m} cannot in general be represented simply by an inverse power law of meteor masses with constant exponent s. The Arietids are found to have $s>2$ over the whole range of magnitudes investigated whereas the Perseids, Quadrantids and Geminids appear to have $s<2$ for the faint and $s \gtrsim 2$ for the bright meteors. We conclude that the greater part of the total meteoric mass is contributed by the following ranges of magnitudes: Arietids, fainter than 7^{m}; Perseids, Quadrantids and Geminids, 2^{m} to 5^{m}. In the case of the Geminids, the concentration of meteoric mass into this magnitude range appears to increase during the course of activity of the shower. Good agreement is found between radio, telescopic and visual data.

---

§ 1. INTRODUCTION

The series of papers of which this forms a part was planned originally by Lovell and Clegg who dealt in Part I (1948) with the mechanism of radio reflections from meteor trails (subsequently extended by Herlofson 1951, and Kaiser and Closs 1952). Part III by Greenhow (1952) discusses the behaviour of the meteor trails subsequent to their formation and the factors which determine the duration of the transient echoes. Part IV (Billam and Browne 1956) discusses the effects of plasma resonance on meteor echoes. The present paper, as Part II of the series, is concerned with the deduction, from the radio echo data, of the distributions of meteor magnitudes and masses.

§ 2. PARAMETERS OF THE MAGNITUDE AND MASS DISTRIBUTIONS

According to the theory of evaporation of meteors, which assumes that the particle makes individual impacts with air molecules, the luminous intensity I and electron line density α at the brightest and most heavily ionized point on a meteor trail are related to the meteor mass m by $I_{\max} \propto Km$, $\alpha_{\max} \propto Jm$ where K and J are the luminous and ionizing efficiencies respectively, assumed to be independent of m. The visual magnitude M_{v} is usually taken to be

$$M_{\mathrm{v}} = \text{const.} - 2{\cdot}5 \log_{10} I_{\max}. \qquad \ldots\ldots(1)$$

Levin (1953, 1955) suggests, however, that the magnitude given by a visual observer may characterize some average luminous intensity, $\bar{I} \propto I_{\max}{}^{x}$, where x is a constant assumed to lie between 0·5 and 1·0. Thus

$$\left.\begin{aligned} M_v &= \text{const.} - 2{\cdot}5x \log_{10} I_{\max} \\ &= \text{const.} - 2{\cdot}5x \log_{10} m. \end{aligned}\right\} \qquad \ldots\ldots(2)$$

The radio echo observations measure the electron line density in the vicinity of the intersection with the meteor trail of the normal from the observer (specular reflection). Following the suggestion of Kaiser (1955) a quantity called the 'radio magnitude' may be defined by

$$\left.\begin{aligned} M_r &= 35 - 2{\cdot}5 \log_{10} \alpha_{\max} \\ &= \text{const.} - 2{\cdot}5 \log_{10} m. \end{aligned}\right\} \qquad \ldots\ldots(3)$$

The numerical constant has been chosen to give the best fit with the visual scale for Perseid meteors (velocity 60 km sec^{-1}).

It should be noted that if K and J are subsequently found to vary with meteor mass, the conclusions in this paper remain unaltered except that they refer to the distribution of the quantities $I_{\max}$ and $\alpha_{\max}$ rather than to the meteor mass.

Following previous practice the frequency distribution of both radio and visual magnitudes will be assumed to have the form

$$\nu_M \propto a^M \qquad \ldots\ldots(4)$$

where $\nu_M dM$ is the incident flux of meteors with magnitudes between M and $M+dM$. From equations (2) and (3) the corresponding flux of meteors with masses between m and $m+dm$ is found to be

$$\nu_m dm \propto m^{-s} dm \qquad \ldots\ldots(5)$$

where

$$s = 1 + 2{\cdot}5x \log_{10} a. \qquad \ldots\ldots(6)$$

$x = 1$ for radio meteors (by definition of radio magnitude) and, according to Levin, $0{\cdot}5 < x < 1$ for visual meteors.

It is obvious that laws of the form (4) and (5) cannot hold over an infinitely wide range of magnitudes (or masses). They do, however, constitute useful definitions of the magnitude and mass distributions in terms of simple parameters a and s provided these latter do not vary too rapidly with M and m.

§ 3. RADIO ECHO METHODS

Equation (3) defines the radio magnitude in terms of the maximum electron line density along the meteor trail. Existing radio echo techniques measure the electron line density in the vicinity of the specular reflection point which, of course, is not in general at the point of maximum ionization. It has been established that when $\alpha < 10^{12}$ cm^{-1} the echo amplitude A is directly proportional to α, and the echo duration τ is proportional to $\lambda^2 p$ where λ is the wavelength and p the pressure in the vicinity of the reflection point (a characteristic value is $\tau \sim 0{\cdot}1$ sec for $\lambda = 8$ m); these are referred to as short duration or decay type echoes. When α is in excess of a few times 10^{12} cm^{-1}, we find $A \propto \alpha^{1/4}$ and $\tau \propto \lambda^2 p$; these are referred to as persistent echoes.

Three kinds of data are available for the determination of radio magnitude distributions.

3.1. *Total Echo Rate versus Equipment Parameters; Echo Amplitude Distribution*

It is assumed that the sensitivity of the equipment is such that the majority of echoes are of the decay type. For a given aerial directional diagram the echo rate can be defined in terms of an equipment sensitivity factor F given by

$$F=(P\lambda^3/\epsilon)^{1/2} \qquad \ldots\ldots(7)$$

where P, λ and ϵ are the transmitted power, wavelength and receiver sensitivity (detectable echo power) respectively. On the assumption of an inverse power law for the mass distribution the echo rates N_1, N_2 corresponding to sensitivity factors F_1, F_2 are related by (Kaiser 1953)

$$N_1/N_2=(F_1/F_2)^{s-1}[1-\tfrac{1}{3}(H/\bar{h})\ln(F_1/F_2)]^{(3s-5)/2} \qquad \ldots\ldots(8)$$

where h is the mean height of the reflection points on the meteor trails and H is the atmospheric scale height. In fact, putting $H\sim 7$ km, $\bar{h}\sim 100$ km, the logarithmic term has a negligible influence on the values of s deduced from (8) and hence the total echo rate is given by

$$N\propto F^{s-1}. \qquad \ldots\ldots(9)$$

The echo amplitude distribution for a given equipment is obtained by putting $\epsilon\propto A^2$, hence the differential echo amplitude distribution is likewise given by an inverse law

$$\nu_A dA\propto A^{-s}dA. \qquad \ldots\ldots(10)$$

Echo amplitudes were measured on a frequency of 55 Mc/s, using a vertically polarized steerable Yagi, which was directed to an azimuth and elevation differing by 180° and 90° respectively from the radiant position of the meteor shower under observation. This geometry ensured that plasma resonance effects did not complicate the echo amplitude measurements (Billam and Browne 1955, 1956). A pulse repetition frequency of 600 per second was used, while the receiver gain was switched between two preset values at half this frequency. In this way, a wide range of echo amplitudes could be measured accurately without producing off-scale deflections on the display. The range–amplitude display is photographed by an automatic recorder triggered by the meteor echoes. The equipment sensitivity was maintained throughout runs by frequent checking of the transmitter power, the receiver noise factor and the recorder triggering level. The amplifier was calibrated and found to be linear for signal amplitudes less than 100 times the r.m.s. noise voltage.

The sensitivity of the equipment can be estimated either from a knowledge of the various parameters of the transmitter, receiver and aerial, or directly from analysis of the observed amplitude distribution. Kaiser's theory summarized above predicts that if the logarithm of the number of echoes N_A of amplitudes greater than A is plotted against $\log A$, the slope will change from $s-1$ to $4(s-1)$ at an amplitude corresponding to a line density which depends only slightly on s, and is given by 5×10^{11} electrons cm^{-1} for $s=2$. A typical plot is shown in figure 1, and it can be seen that its general form is in good agreement with prediction. The two methods of estimating equipment sensitivity agree to within 15%, and show that echoes of amplitude 3 × r.m.s. noise voltage are given by meteor trails of $\alpha=6\times 10^{10}$ electrons cm^{-1} at a range of 200 km (radio magnitude $=8{\cdot}1^{\mathrm{m}}$ at 200 km).

For the analysis echoes were selected on a range basis, to exclude meteors appearing in the side lobes, while echoes from trails twisted into the specular condition by turbulence were eliminated by selecting only echoes showing the Fresnel velocity–diffraction pattern. The sporadic meteor rate at the time of each shower was measured and an appropriate correction made to the observed amplitude distribution.

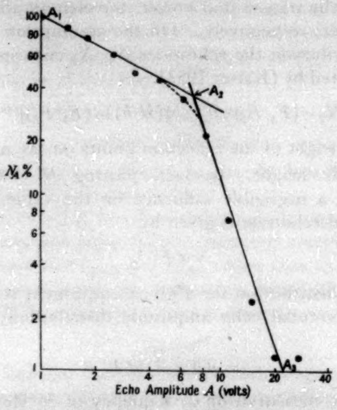


Figure 1. Typical double-logarithmic plot of N_A, the number of echoes of amplitudes exceeding A, against A. The measurements were made during the 1954 Perseid shower. The portion of the curve A_1A_2 has a predicted slope $s-1$, while A_2A_3 has a predicted slope of $4(s-1)$. A_2 indicates the echo amplitude corresponding to trails of density 5×10^{11} electron cm^{-1}.

3.2. *Radio Echo Height Distribution*

A theory of the meteor height distribution obtained from radio echo observations has been developed by Kaiser (1954) who has shown that the width of the height distribution for a homogeneous velocity group of meteors depends significantly only on the atmospheric scale height and the mass distribution of the incident particles. The r.m.s. deviation, δh, from the mean height has been evaluated as a function of the exponent s (δh increases for decreasing s). The theory has been applied by Evans (1954) who has taken the value $s = 2{\cdot}0$ for sporadic meteors (see § 4) and deduced the atmospheric scale height as a function of altitude (over the range 85–105 km) from measurements of the height distribution of meteors in selected velocity intervals.

A major meteor shower itself comprises a homogeneous velocity group, hence no velocity selection is necessary and a satisfactory height distribution can be obtained from a few hours' observations. The danger still exists that extreme observations, due to sporadic meteors with different velocities from that of the shower, may influence seriously the determination of δh. We have therefore measured the difference between the heights at which the echo frequency is equal to one half its maximum value and applied a correction to obtain δh. An additional small correction is applied to take account of the estimated height scatter due to experimental errors.

The mass distribution for the various showers has been deduced from the theory, using the scale height values of Evans. Thus the values of δh are compared with those for sporadic meteors of the same velocity, taking $s = 2{\cdot}0$ for the latter.

Meteor heights have been measured on two equipments which operate at frequencies of 69 and 36·6 Mc/s respectively. The former, which has been described by Clegg and Davidson (1950) radiates pulses of 50 kw peak power from a dipole-plus-reflector aerial. Echo signals are received alternately on similar aerials at heights $0{\cdot}5\lambda$ and $0{\cdot}75\lambda$ above a level reflecting surface and the angle of elevation of the reflection points calculated from the ratio of the two signals. This, with the range, enables the height of the reflection point to be determined. The 36·6 Mc/s equipment is similar, except that the receiving aerials are at heights $0{\cdot}25\lambda$ and $0{\cdot}5\lambda$, are portable, and hence may be set up to face different azimuths (for measurements on meteor showers with different radiant declinations).

The accuracy of a single measurement on either equipment lies between 1 and 2 km and their sensitivities are such that the smallest detectable electron line density in a meteor trail is of the order $2 \times 10^{11}\,\mathrm{cm}^{-1}$ (i.e. $M_r = 6{\cdot}8$). The observations therefore relate mainly to meteors in the range 5^m to 7^m.

It has been noted (§ 3) that $\tau \propto \lambda^2 p$ for the decay type echoes. Thus at great heights the rapid radial diffusion of the ionization may result in a failure to record echoes. This limitation is most serious for the short wavelength (69 Mc/s) equipment and is noticeable at 100–105 km height. On 36 Mc/s the duration at a given height will be longer by a factor of 4 and the results should be reliable up to about 110 km.

3.3. *Distribution of Persistent Echo Durations*

According to Kaiser (1953, 1955), the predicted maximum echo duration τ_{max} which can be produced by meteors of given velocity and angle of incidence is related to α_{max} by

$$\tau_{\mathrm{max}} \propto \alpha_{\mathrm{max}}{}^{4/3}\lambda^2 \qquad \ldots\ldots(11\,a)$$

where λ is the radio wavelength. From this it follows that the differential distribution of persistent echo durations is given by

$$\nu_\tau d\tau \propto \tau^{-(3s+1)/4} d\tau. \qquad \ldots\ldots(12\,a)$$

However, Billam and Browne (1956) have shown that measurements of echo amplitude and duration for meteors of line density greater than 10^{12} electrons cm^{-1} suggest a relation of the form

$$\tau_{\mathrm{max}} \propto \alpha^{0{\cdot}84}\lambda^2. \qquad \ldots\ldots(11\,b)$$

If this relation rather than the predicted (11 *a*) be used, the differential distribution of persistent echo durations is modified:

$$\nu_\tau d\tau \propto \tau^{-0{\cdot}2(6s-1)} d\tau. \qquad \ldots\ldots(12\,b)$$

In this paper (12 *b*) has been used to obtain values of *s* from observed duration distributions, but it should be remembered that the predicted and observed relations between echo durations and amplitudes are not in agreement, and that there is some ambiguity in the results.

The theory of duration distributions makes the assumption that all echoes are caused by meteor trails formed in the specularly reflecting condition. Browne (unpublished) has considered the effect of 'drifting' trails—the twisting of meteor trails by winds after formation—and shown that, in general, the observed duration distribution will yield values of s lower than the correct value unless only those echoes are selected which show the Fresnel velocity–diffraction pattern and so arise at trails formed in the specular condition. Unless a range–amplitude display is used as well as the usual range–duration display, it is difficult to select echoes on this basis, so that values of s obtained from duration distributions are likely to be too small unless the echoes analysed have been very carefully selected. This is illustrated in figure 2 where differential duration distributions obtained during 1954 Perseids are shown for (*a*) all echoes observed, (*b*) selected echoes, reaching maximum amplitude within 0·1 sec, with amplitude never exceeding double the initial maximum. The values of s obtained are, respectively, 1·1 and 1·8.

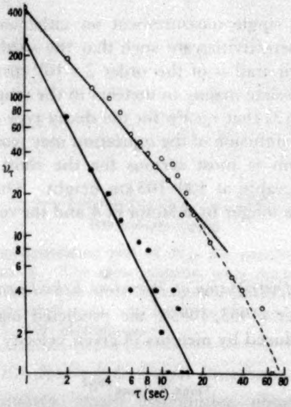


Figure 2. Differential duration distribution at 36 Mc/s during the 1954 Perseid shower. The plot is double-logarithmic, showing the number of echoes of duration between τ and $\tau+1$ sec. The solid circles represent echoes specially selected to exclude those due to trails distorted by wind, while the open circles represent all the echoes observed. The slopes of the two lines are 1·95 and 1·17 respectively, leading to values of s of 1·8 and 1·1.

The duration distributions given in this paper were obtained with three equipments: (i) The meteor radiant survey apparatus (Aspinall *et al.* 1951) operating on a frequency of 72 Mc/s. The echo durations are recorded on a range–time display with continuously moving film. Echoes have not been selected to exclude meteor trails 'drifting' into a reflecting condition, so that the values of s obtained with this equipment are probably under-estimates. (ii) The meteor polarization equipment (Billam and Browne 1955, see also § 3.1 above) operating on a frequency of 55 Mc/s. Echo durations on a time–amplitude display are measured to the nearest second by counting the number of frames on which each echo is recorded, the recorder triggering rate being set to one frame per second. Echoes were selected as described above. (iii) A special apparatus operating at 36 Mc/s, using a horizontally directed Yagi aerial. Echoes are recorded on continuously moving film in a display which includes range–time and amplitude–time tubes, as well as one second marker pulses. Echoes are selected for analysis as described above. The durations in general relate to meteors brighter than $+3^m$.

§ 4. MAGNITUDE AND MASS DISTRIBUTIONS FOR SPORADIC METEORS

The theory of § 3.1 has been applied by Kaiser (1953, 1954) to the experimental values of McKinley (1951) who found $N_1/N_2=2{\cdot}53$, $3{\cdot}69$, 45 respectively for values $F_1/F_2=2{\cdot}5$, $4{\cdot}0$, 43 (covering the magnitude range 5^m to 10^m). The values obtained for s are 2·02, 1·97, 2·02 and the value adopted is $s=2{\cdot}00\pm0{\cdot}02$, $a=2{\cdot}51\pm0{\cdot}05$ for $10^m>M>5^m$. Using the equipment described in § 3.1 and plotting logarithmically the number of echoes N_A exceeding amplitude A against A, we have found for 140 echoes on 20th–21st, 21st–22nd August 1954: $s=1{\cdot}98\pm0{\cdot}02$, $a=2{\cdot}47\pm0{\cdot}05$ for $8^m>M>5^m$ and $s=2{\cdot}4\pm0{\cdot}6$, $a=3{\cdot}6\pm1{\cdot}4$ for $5^m>M>0^m$. The greater probable error for the brighter meteors is due to their relative scarcity.

The method of § 3.2 can be used only to compare the magnitude and mass distribution of different groups of meteors, unless a value is assumed for the atmospheric scale height. In fact, the measured height distributions for selected velocity groups of sporadic meteors have been used, together with the mass distribution obtained above, to determine the scale height (Evans 1954). The agreement obtained with rocket measurements may be accepted as confirmation of the value of s given above, and also of the assumption that the value of s is substantially the same for the selected velocity groups between 20 and 60 km sec^{-1}.

The method of § 3.3 which uses the radio echo duration distributions has been applied to measurements by McKinley (1954) on a wavelength of 9·22 m during June and August 1950. The results for non-shower periods during these months lead to a value of s of 1·61 (using formula (12*b*)) for the magnitude range 2^m to 4^m, but these echoes were not selected to exclude 'drifting' meteor trails, so that this value of s represents a lower limit.

The agreement between the values of s found for different ranges of radio magnitudes and from independent data is fair, and it seems reasonable to adopt the values $s=2{\cdot}0$, $a=2{\cdot}5$ for radio meteor magnitudes between 2^m and 10^m.

§ 5. MAGNITUDE AND MASS DISTRIBUTION FOR SHOWER METEORS

5.1. *Quadrantids*

A Quadrantid height distribution obtained on 2nd–3rd January 1953 on 36 Mc/s is shown in figure 3. Treatment of these data as described in § 3.2 gives $\delta h=8{\cdot}45$ km and $s=1{\cdot}78\pm0{\cdot}06$, $a=2{\cdot}05\pm0{\cdot}11$. The distribution of echo durations for the Quadrantids obtained with the 72 Mc/s equipment is given in figure 4 and suggests the values $s=1{\cdot}67$, $a=1{\cdot}85$, but this is certainly an under-estimate, for the reasons given in § 3.3.

The period of the shower activity is short and the radiant equipment observations are limited to approximately 1½ hours near transit; it is thus not possible to determine whether there is a systematic variation in the mass distribution during the period of activity. On 3rd January 1953, however, when the echo rate rose to several times that normally experienced, the well-defined subcentre of the radiant contained a relatively greater proportion of small meteors than the surrounding radiant area (Bullough 1954).

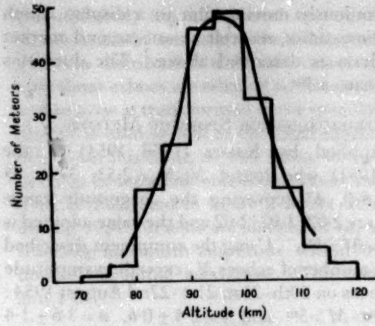


Figure 3. Height distributions, Quadrantids, 2nd–3rd January 1953 (36 Mc/s).

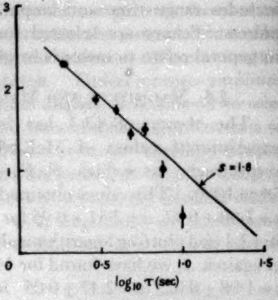


Figure 4. Echo duration distributions, Quadrantids 1951–52–53 (72 Mc/s). The echoes used were not selected to exclude ‘ drifting ’ echoes.

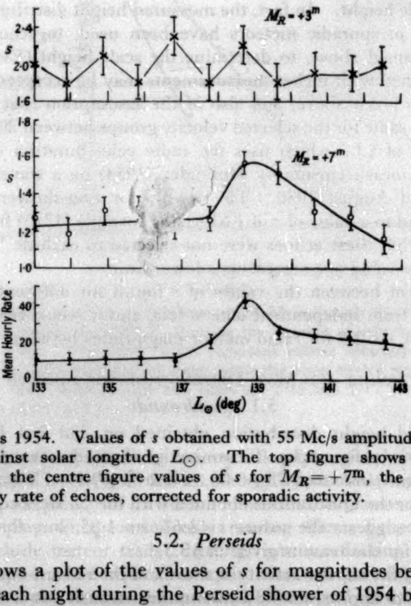


Figure 5. Perseids 1954. Values of *s* obtained with 55 Mc/s amplitude analysis method, plotted against solar longitude $L_\odot$. The top figure shows values of *s* from $M_R = +3^m$, the centre figure values of *s* for $M_R = +7^m$, the bottom figure the mean hourly rate of echoes, corrected for sporadic activity.

5.2. *Perseids*

Figure 5 shows a plot of the values of *s* for magnitudes between $+8^m$ and $+6^m$ obtained each night during the Perseid shower of 1954 by the method of § 3.1, using the 55 Mc/s polarization equipment. It can be seen that for this magnitude range, *s* reaches a maximum of $1{\cdot}55 \pm 0{\cdot}05$ at the shower maximum. Using the same method for brighter meteors, a mean value of $s = 1{\cdot}96 \pm 0{\cdot}17$ is obtained for magnitudes between $+4^m$ and 0^m, while there is no detectable variation of this value with time.

Figures 6 (*a*) and (*b*) give the heights of meteors observed on 36 Mc/s during the periods 8th–10th August 1952 and 10th–14th August 1953 respectively. The difference between the two distributions is not great, although figure 6 (*a*), which relates to the earlier part of the shower activity, has a noticeably flat top which may be due to a deficiency in faint meteors (relative to the numbers predicted by an inverse power law of meteor masses). From figure 6 (*b*) the mean height is found to be 102·5 km and $\delta h = 12{\cdot}5$ km. This corresponds to a mean value of $s = 1{\cdot}59 \pm 0{\cdot}03$ over the period 10th–14th August in the magnitude range $+5^{\mathrm{m}}$ to $+7^{\mathrm{m}}$.

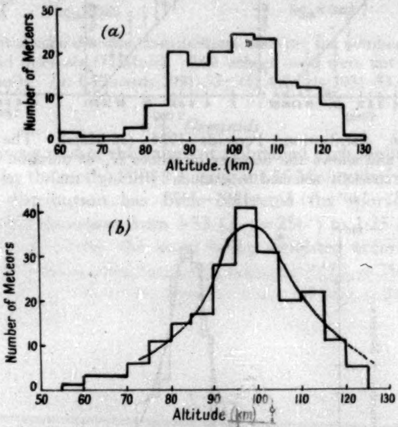


Figure 6. Height distributions, Perseids (36 Mc/s). (*a*) 8th–10th August 1952; (*b*) 10th–14th August 1953.

Duration measurements obtained with the 36 Mc/s equipment are shown in figure 7, which shows the number of echoes exceeding a certain duration plotted against the duration. The echoes have been selected as described in § 3.3 to exclude meteor trails which have 'drifted' in. The values of *s* obtained from these plots refer to the magnitude range $+3^{\mathrm{m}}$ to 0^{m}, and show insignificant variation with solar longitude, giving a mean value of $s = 2{\cdot}00 \pm 0{\cdot}02$, in good agreement with the result obtained from the echo amplitude analysis. The grouping of durations between 20 seconds and 50 seconds (about $+1^{\mathrm{m}}$) for 11th–12th–13th August is remarkable. A more detailed analysis shows it to be spread evenly over 11th–12th and 12th–13th August. The grouping also appears in the duration distributions obtained with the 72 Mc/s equipment between durations of 5 seconds and 10 seconds, again corresponding to $+1^{\mathrm{m}}$ (duration $\propto$ wavelength^2). The 72 Mc/s durations, however, were not selected to exclude 'drifting' echoes, and are averaged over the whole Perseid period, so cannot be used to obtain a reliable value of *s*.

5.3. *Arietids*

Figure 8 gives the results of height measurements made on 69 Mc/s during 3rd–9th June 1952 and 1953. The echoes have been divided into 2 km height groups (instead of the 5 km grouping used for the other showers), and the extreme narrowness of the height distribution for this summer daytime meteor shower is at once apparent. The mean height is 92·5 km and $\delta h = 4{\cdot}7$ km corresponding to $s = 2{\cdot}5$, $a = 4{\cdot}0$.

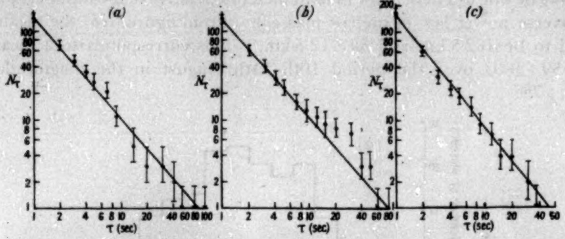


Figure 7. Duration distribution, Perseids 1954 (36 Mc/s). The plot is double-logarithmic, and shows the number of echoes N_τ of duration exceeding τ. The echoes were specially selected to exclude 'drifting' trails. (a) $L_\odot = 135° - 138°$; (b) $L_\odot = 138° - 140°$; (c) $L_\odot = 140° - 142°$.

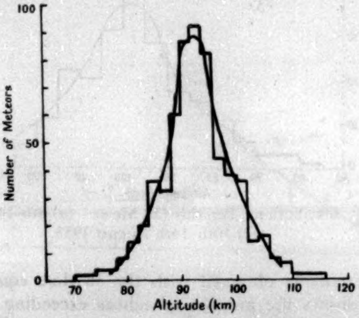


Figure 8. Height distributions, Arietids, 3rd–9th June 1952–53 (69 Mc/s).

In actual fact the observations made using the height-finding equipment with its low-gain aerials will refer to a mixture of Arietid and ζ-Perseid meteors, since both showers are active at the periods concerned and their radiants are separated by only one hour in right ascension. The velocities are 29 and 38 km sec^{-1} for the ζ-Perseids and Arietids, respectively (Davies and Greenhow 1951), hence according to Evans (1954) the difference in mean heights of the two showers will be about 3 km. If we assume the ζ-Perseids to have a similar mass distribution to the Arietids (see below) but about one half of the hourly rate, we obtain the corrected value $\delta h = 4{\cdot}3 \pm 0{\cdot}4$ km and hence $s = 2{\cdot}7 \pm 0{\cdot}3$, $a = 5{\cdot}0 \pm 1{\cdot}4$.

The duration distributions in figures 9 (a) and (b) were obtained on 72 Mc/s and refer to the ζ-Perseid and Arietid showers respectively, it being possible to resolve these two showers with the radiant equipment. For the Arietids the values $s = 1{\cdot}8$, $a = 2{\cdot}1$ were obtained, while for the ζ-Perseids the values $s = 2{\cdot}2$, $a = 3{\cdot}0$. The durations were not specially selected, so that the values of s and a represent lower limits, for the reasons discussed in § 3.3.

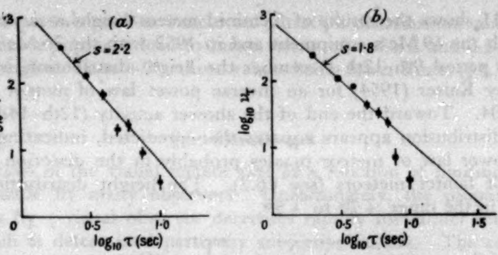


Figure 9. Differential echo duration distributions, showing the number of echoes observed in 1·5 second intervals (72 Mc/s). The echoes used were not selected to exclude 'drifting' trails. (*a*) ζ-Perseids 1951–53; (*b*) Arietids 1951–53.

5.4. *Geminids*

Figure 10 shows a plot of the values of *s* obtained each night during the 1954 Geminid shower by the method of § 3.1 using the 55 Mc/s polarization equipment. The amplitude distribution has been corrected for sporadic meteors. At $M=+7^{\mathrm{m}}$, *s* steadily decreases from 1·53 $(L_{\odot}=256^{\circ})$ to 1·25 $(L_{\odot}=263^{\circ})$ with a mean value of 1·45 ± 0·01, the mean being weighted according to the echo rate. At $M=+3^{\mathrm{m}}$, on the other hand, *s* rises from 1·7 $(L_{\odot}=256^{\circ})$ to a maximum of $s=2{\cdot}7$ $(L_{\odot}=259^{\circ})$ and then falls again to reach 1·7 $(L_{\odot}=263^{\circ})$, the weighted mean value being 2·24 ± 0·04.

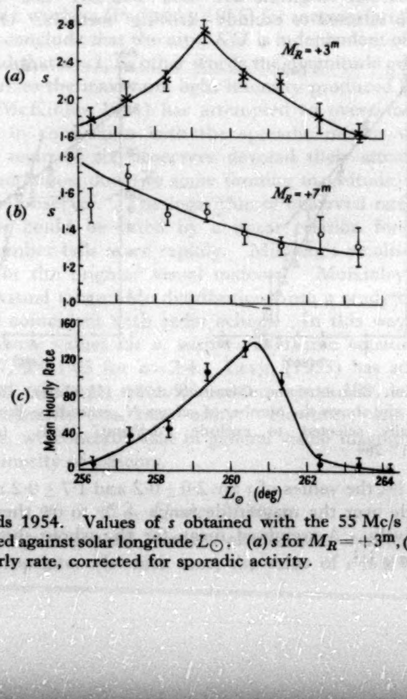


Figure 10. Geminids 1954. Values of *s* obtained with the 55 Mc/s amplitude analysis method, plotted against solar longitude $L_{\odot}$. (*a*) *s* for $M_R=+3^{\mathrm{m}}$, (*b*) *s* for $M_R=+7^{\mathrm{m}}$; (*c*) Mean hourly rate, corrected for sporadic activity.

Figure 11 shows the results of Geminid meteor height measurements made in 1951 with the 69 Mc/s equipment and in 1952 with the 36 Mc/s equipment. During the period 9th–12th December the height distribution is of the form predicted by Kaiser (1954) for an inverse power law of meteor masses, with $s = 1{\cdot}62 \pm 0{\cdot}04$. Toward the end of the shower activity (12th–14th December) the height distribution appears squarer than predicted, indicating a departure from the power law of meteor masses probably in the direction of a relative deficiency of fainter meteors (see § 6.2). The height distributions apply to magnitudes in the range $+7^{\mathrm{m}}$ to $+5^{\mathrm{m}}$.

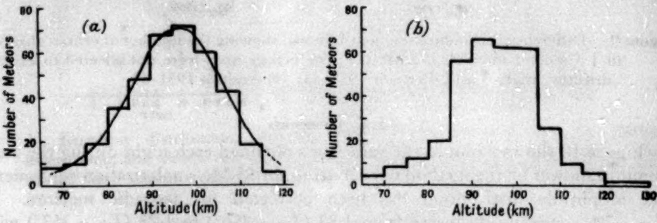


Figure 11. Height distributions, Geminids, 1951 (69 Mc/s), 1952 (36 Mc/s). (*a*) 9th–12th December; (*b*) 12th–14th December.

Geminid echo duration distributions obtained with the 55 Mc/s polarization equipment are shown in figures 12 (*a*) and (*b*), for the periods of shower activity corresponding to solar longitude 256°–260° and 261°–263° respectively. The echoes have been selected to exclude 'drifting' meteors. Over the magnitude range $+3^{\mathrm{m}}$ to $+1^{\mathrm{m}}$, the values of s are $2{\cdot}0 \pm 0{\cdot}2$ and $1{\cdot}7 \pm 0{\cdot}2$ respectively for the two periods, while over the magnitude range $+2^{\mathrm{m}}$ to 0^{m} they are $2{\cdot}6 \pm 0{\cdot}2$ and $3{\cdot}7 \pm 0{\cdot}3$ respectively. A weighted mean for the whole shower for $M < 3^{\mathrm{m}}$ is given by $s = 2{\cdot}3$, $a = 3{\cdot}3$.

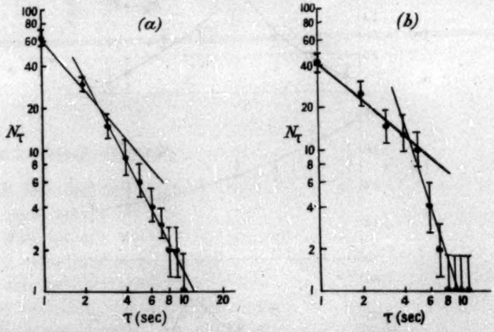


Figure 12. Duration distributions, Geminids 1954 (55 Mc/s). The plot is double-logarithmic, and shows the number of echoes N_τ exceeding duration τ. The echoes were specially selected to exclude 'drifting' trails. (*a*) $L_\odot = 256°–260°$; (*b*) $L_\odot = 261°–264°$.

For the same two periods unselected duration distributions made with the 72 Mc/s radiant equipment lead to mean values of s of 2·2 and 2·5 respectively for meteors brighter than $+3^m$, these values being lower limits (§3.3).

§6. DISCUSSION

Observations of the visual meteor rate as a function of limiting magnitude have been made by many observers. Unfortunately, the proportion of the meteors seen by a visual observer decreases rapidly for fainter meteors by an amount which is determined partly by subjective factors. The conversion of the observed into the true meteor distribution thus involves large corrections for the fainter meteors ($M > +3^m$) (Öpik 1922 a, Ceplecha 1950). The situation may be better for telescopic observations since the field of view is well defined. The radio measurements do not suffer such defects, but it should be noted that strictly speaking they indicate distributions of electron line densities, while the visual measurements indicate distributions of luminosity. Kaiser (1955) has indicated how measurements of meteor rate and mass distribution can be used to obtain values of the true incident meteor flux per unit area.

6.1. *Sporadic Meteors*

The radio echo measurements indicate that an inverse square differential mass law is followed over the range 2^m to 10^m. This corresponds to an increase of 2·5 times in frequency for each fainter meteor magnitude, which is precisely the result obtained by Watson (1941) for telescopic meteors down to magnitude $+10^m$ and by Soviet workers (Fedynsky 1955) for meteors brighter than $+11^m$. We are thus led to conclude that the ratio K/J is independent of magnitude over the above range and that $x = 1$, in other words the magnitude given by telescopic observers characterizes the maximum light intensity produced by a meteor.

Millman (see McKinley 1951) has attempted to overcome the difficulties referred to above in connection with the sporadic meteor visual magnitude distribution. His team of six observers devoted their attention to different areas of the sky and hence, down to some limiting magnitude, all meteors were seen by at least one observer. The logarithm of observed rate as a function of limiting magnitude could be fitted by a linear relation for $-3 < M_v < +2$; beyond this the number falls more rapidly. Millman's results obtained in this way give $a = 2{\cdot}7$ for the brighter visual meteors. McKinley (1951) has also deduced the true visual magnitude distribution from a study of the fraction of visual occurrences coincident with radio echoes. In this way he gets $a = 2{\cdot}4$. Substituting the above values for a, with $s = 2{\cdot}00$, into equation (6) we obtain $x = 0{\cdot}93$ for $a = 2{\cdot}7$, $x = 1{\cdot}05$ for $a = 2{\cdot}4$. Levin (1955) has adopted the value $a = 3$ for visual and telescopic sporadic meteors up to 10^m, leading to a value $x = 0{\cdot}83$, i.e. significantly less than unity. However, using Millman and McKinley's results, we conclude that in general visual magnitudes characterize the maximum luminosity of meteors.

6.2. *Shower Meteors*

Table 1 exhibits the general features of the radio results and also visual results, where appropriate. In each case, the value of s is a weighted average

Table 1. Values of *s* for Shower Meteors

(corresponding values of *a* are given by $\log_{10} a = (s-1)/2{\cdot}5$)

Shower	(1) $M_R = +7$	(2) $+7 > M_R > +5$	(3) $M_R = +3$	(4) $+2 > M_R > 0$	(5) $M_v < +3$
Quadrantids	—	1·78	—	1·8	2·5[a]
Perseids	1·38	1·59	2·0	2·0	2·1[b]
Arietids	—	2·7	—	1·8†	—
ζ-Perseids	—	—	—	2·2†	—
Geminids	1·45	1·62	2·24	2·3	2·3[a]

(1) Echo amplitude method; (2) echo height method; (3) echo amplitude method; (4) echo duration method; (5) visual observation.

† Results obtained from unselected durations, and so representing only a lower limit to *s* or *a* (cf. § 3.3).

[a] Visual results after Levin (1955).

[b] Mean value from analysis of the visual results of Öpik 1922 b, Levin 1955, Kresak and Vozarova 1953, Hruška 1954, Ceplecha 1952.

for the whole shower. The faint meteors ($7 > M_R > 5$) studied by the amplitude and height methods have, with the exception of the Arietids, $s < 2$. Thus for these showers (Quadrantids, Perseids and Geminids) the increase in frequency with increasing magnitude is less rapid than is the case for sporadic meteors. For the brighter meteors in all the showers observed with the amplitude or duration methods, $s \gtrsim 2$, suggesting that for the showers as a whole there is a departure from inverse power of meteor masses, the rate of fall off in numbers with increasing mass becoming greater for the brighter meteors.

The Arietids and ζ-Perseids seem to be exceptional in having a large value of *s* for both faint and bright meteors.

It can be seen that where visual observations are available, they are in good agreement with radio observations. By comparing the radio and visual values of *s* for meteors brighter than $+3^m$, we can obtain values of the constant *x* (cf. equation (6)). This can be done for the Perseid and Geminid showers, and the results are shown in table 2. The Quadrantids have not been included because the value of *s* obtained for the brighter meteors by the radio duration method is unreliable. It has already been shown that, for sporadic meteors, $x = 1$ for Millman's measurements of *s* for visual meteors, but $x = 0{\cdot}83$ for Levin's measurements on visual meteors. Because of the errors involved, it would seem reasonable to adopt the value $x = 1{\cdot}0$.

Table 2. Comparison of Visual and Radio Values of *s* to find *x*

Shower	Radio $s : s_r$	Visual $s : s_v$	$x = (s_r - 1)/(s_v - 1)$
Perseids	2·00	2·1	0·91
Geminids	2·3	2·3	1·00

In view of the considerable amount of data available, the Perseid and Geminid showers will be discussed in more detail.

Perseids. In general, the mean value of *s* increases from about 1·4 for meteors of $+7^m$ to 2·1 for meteors of 0^m. This indicates that the total mass per magnitude range is concentrated around magnitudes $+5^m$ to $+3^m$. It is also interesting to note from figure 5 that when the shower rate is a maximum, the number of faint meteors ($M_R > 6^m$) rises relative to that of the brighter meteors, falling again towards the end of the shower.

Geminids. Like the Perseid meteors, Geminid meteors show a pronounced increase in s from $+7^{\mathrm{m}}$ to 0^{m}. The changes of s during the shower indicating that not only do the meteors become progressively concentrated during the shower into a small magnitude range, but that the magnitude at which the meteors are most concentrated decreases from about $+4^{\mathrm{m}}$ to about $+2^{\mathrm{m}}$.

ACKNOWLEDGMENTS

The authors were fortunate in having access to a pre-publication copy of the paper on visual magnitude distributions by Dr. B. J. Levin. We are indebted to Professor A. C. B. Lovell for providing facilities for the work and for reading and criticizing the manuscript. We wish also to thank Mr. E. R. Billam for assistance in carrying out the observations at 55 Mc/s. We acknowledge the financial support towards the construction and maintenance of equipment given by the Department of Scientific and Industrial Research to whom one of us (S. E.) is also indebted for the award of a maintenance grant.

REFERENCES

ASPINALL, A., CLEGG, J. A., and HAWKINS, G. S., 1951, *Phil. Mag.*, **42**, 504.
BILLAM, E. R., and BROWNE, I. C., 1955, *Meteors* (edited by T. R. Kaiser) (London: Pergamon Press), p. 73; 1956, *Proc. Phys. Soc.* B, **69**, 98.
BULLOUGH, K., 1954, *Jodrell Bank Ann.*, **1**, 37.
CEPLECHA, Z., 1950, *Bull. Astr. Insts. Csl.*, **2**, 146; 1952, *Ibid.*, **3**, 65.
CLEGG, J. A., and DAVIDSON., I. A., 1950, *Phil. Mag.*, **41**, 77.
DAVIES, J. G., and GREENHOW, J. S., 1951, *Mon. Not. Roy. Astr. Soc.*, **111**, 26.
EVANS, S., 1954, *Mon. Not. Roy. Astr. Soc.*, **114**, 63.
FEDYNSKY, V. V., 1955, *Meteors* (edited by T. R. Kaiser) (London: Pergamon Press), p. 188.
GREENHOW, J. S., 1952, *Proc. Phys. Soc.* B, **65**, 169.
HERLOFSON, N., 1951, *Ark. Fys.*, **3**, 247.
HRUŠKA, A., 1954, *Bull. Astr. Insts. Csl.*, **5**, 13.
KAISER, T. R., 1953, *Advanc. Phys.*, **2**, 495; 1954, *Mon. Not. Roy. Astr. Soc.*, **114**, 39, 52; 1955, *Meteors* (edited by T. R. Kaiser) (London: Pergamon Press), p. 119.
KAISER, T. R., and CLOSS, R. L., 1952, *Phil. Mag.*, **43**, 1.
KRESAK, L., and VOZAROVA, M., 1953, *Bull. Astr. Insts. Csl.*, **4**, 139.
LEVIN, B. J., 1953, *Doklady Acad. Sci. U.S.S.R.*, **90**, 513; 1955, *Meteors* (edited by T. R. Kaiser) (London: Pergamon Press), p. 131.
LOVELL, A. C. B., and CLEGG, J. A., 1948, *Proc. Phys. Soc.*, **60**, 491.
MCKINLEY, D. W. R., 1951, *Canad. J. Phys.*, **29**, 403; 1954, *Ibid.*, **32**, 450.
ÖPIK, E. J., 1922 a, *Publ. Univ. Sternw. Darpat*, **25**, No. 1; 1922 b, *Ibid.*, **25**, No. 4 (quoted by A. C. B. Lovell, *Meteor Astronomy* (Oxford: Clarendon Press), 1954, p. 384).
WATSON, F., 1941, *Between the Planets* (Philadelphia: Blakiston).

Characteristics of Radio Echoes from Meteor Trails
IV: Polarization Effects†

By E. R. BILLAM and I. C. BROWNE
Jodrell Bank Experimental Station, University of Manchester

Communicated by A. C. B. Lovell; MS. received 1st September 1955

Abstract. Observations have been made of radio echoes from meteor trails at 55 Mc/s in order to test the theoretical estimates by Kaiser and Closs of plasma resonance effects. The aerial array consisted of two Yagis mounted with mutually perpendicular planes of polarization on a common axis. The polarization effects observed are in good agreement with prediction for both short and long duration echoes, but some unexpected results were obtained for meteor trails with line densities in the transition region of 10^{12} electrons cm^{-1}. Some calculations have been made of the effect of the finite meteor velocity upon observed polarization effects, and are in good agreement with observation.

In addition to the polarization measurements, the relationship between the amplitude A and duration T of long duration echoes was investigated. An empirical relation of the form $A \propto T^{0.3}$ was found, in contrast to the predicted law $A \propto T^{3/16}$. It is suggested that the discrepency is a result of diffusion by turbulence in the atmosphere.

---

§ 1. INTRODUCTION

When a meteor passes into the atmosphere it is heated by the impact of air molecules. If the temperature rises sufficiently atoms evaporate from the surface, and, since they are moving with the velocity of the meteor, they have enough kinetic energy to become ionized by individual collisions with air molecules. Thus a column of ionization which can scatter radio waves is formed in the wake of the meteor (see Herlofson 1948 and Kaiser 1953 for full reviews). The ionized trails studied by radio echo methods generally occur between heights of 85 and 105 km, where the mean free path is of the order of a few centimetres.

A comprehensive theory of the scattering of radio waves by meteor trails ('meteor echoes') has been given by Kaiser and Closs (1952), who assumed that the ionization is circularly symmetrical about the axis of the trail, with an electron density $n(r)$ at radius r and time t given by

$$n(r) = n(0) \exp\left[-(r/r_0)^s\right] \qquad \ldots\ldots(1)$$

where $n(0) = \alpha/\pi r_0^2$ and $r_0^2 = 4Dt$; α is the number of electrons per unit length of the trail (the 'line density') and D the diffusion coefficient. The case $s=2$ is of special interest, as it represents a column of ionization diffusing outwards at constant temperature (Huxley 1952). Kaiser and Closs, following Greenhow (1952 a), show that meteor echoes can be divided broadly into two classes, 'long duration' and 'short duration' echoes.

---

† This paper forms Part IV of a series. Part I (Lovell and Clegg 1948) and Part III (Greenhow 1952 a) discuss the intensities of echoes from meteor trails, while Part II (Browne, Bullough, Evans and Kaiser 1956) discusses meteor mass distributions.

(i) *Long duration echoes* occur when the incident radio wave cannot penetrate the column of ionization, which is said to be over-dense. This happens when the condition $|r\sqrt{\kappa}| \geqslant \lambda/2\pi$ is satisfied throughout the trail, where κ is the dielectric constant at radius r and λ is the radio wavelength; this reduces to $\alpha \geqslant 10^{12}$ electrons cm^{-1}. It should be noted that this condition is independent of wavelength. The radio wave can be considered as undergoing reflection at the critical radius where the dielectric constant is zero ('metallic cylinder reflection'), so that as the trail expands by diffusion, the echo amplitude rises, and then falls rapidly as the dielectric constant at the axis of the trail rises to zero. For these over-dense trails it is predicted that the echo amplitude should be proportional to the fourth root of the line density α, while the echo duration τ should be directly proportional to α. The reflection coefficient should be independent of the polarization of the incident wave, except initially when the radius of the trail is less than $\lambda/2\pi$, when it is predicted that the reflection coefficient should be greater for longitudinally polarized waves, i.e. waves with the electric vector parallel to the meteor trail.

(ii) *Short duration echoes* occur when the radio wave penetrates the column of ionization, which is said to be under-dense ($\alpha \lesssim 10^{12}$ electrons cm^{-1}). The echo amplitude decreases exponentially as the trail expands, because of the phase differences across the trail, the duration of the echo being independent of α. The reflection coefficient should be dependent on the polarization of the incident wave, since for transversely polarized waves (with electric vector perpendicular to the axis of the trail) space charge separation will occur, with the possibility of plasma resonance. There will be no such effect for longitudinally polarized waves for which the amplitude of the echo should be proportional to α. Kaiser and Closs have made a detailed study of the polarization effects expected and reach the following conclusions.

(*a*) As the trail expands by diffusion the axial dielectric constant increases from its initial highly negative value towards $+1{\cdot}0$, and for the case $s=2$ plasma resonance will occur as it passes through the value $-1{\cdot}4$. Since the dielectric constant decreases with increasing wavelength it follows that the resonance will occur later and will last longer as the radio wavelength λ is increased.

(*b*) Near plasma resonance there is an increase in the *polarization ratio* ρ defined as the ratio of the amplitude reflection coefficients for transverse and longitudinal polarization respectively. The increase is mainly limited by collision damping, which becomes important in the annular region where the dielectric constant passes through zero and where the electric field becomes infinite because of the condition of continuous electric displacement throughout the trail. It follows that the damping increases with the width of this region and is greater when the radial gradient of dielectric constant is small; that is, the polarization ratio increases with s. For $s=2$, the maximum predicted polarization ratio is $2{\cdot}0$. As the trail expands beyond the resonant condition the polarization ratio falls to unity.

(*c*) Initially, the phase of the transverse reflection coefficient leads that of the longitudinal reflection coefficient by 180°, because before resonance the natural frequency of the trail is greater than the applied radio frequency. As the trail expands the phase lead of the transverse reflection coefficient falls to zero with the plasma frequency.

It has been assumed above that the part of the trail which makes the major contribution to the echo is uniform, i.e. that the time taken for the meteor to cross the first few Fresnel zones is small compared with the duration of the echo. The effect of the finite velocity of the meteor is considered in § 4.3.

Experimental studies of radio reflections from meteor trails have been made by Greenhow (1952 a), who established quantitatively the classification into long and short duration echoes, but did not attempt to study polarization effects. Later Clôss, Clegg and Kaiser (1953) attempted to measure polarization ratios at a wavelength of 4 m, with linearly polarized radio waves, and using a broadside array of dipoles and complementary slots. Their results show fair agreement with theory, but the observations were unsatisfactory because the gain of the slot array was unexpectedly much less than that of the dipole array, while at the same time it was difficult to check the performance of the two aerials. More recently Van Valkenburg (1954) has studied polarization effects in short duration echoes at a wavelength of 13·1 m using helical aerials giving circular polarization. His results do not disagree with prediction but the measurements suffer from two disadvantages. First, it was not possible to identify separately the two reflection coefficients. Second, the maximum polarization ratio as defined above was not measured directly, but a quantity was derived from the measurements which depends on the relative phase as well as on the magnitude of the reflection coefficients. The value of this quantity which corresponded approximately to the maximum of the polarization ratio was obtained for each echo. It can be shown that this represents an upper limit to the maximum polarization ratio, but since the relative phase between the reflection coefficients was not measured it is difficult to establish a close correspondence between Van Valkenburg's results and the predictions of Kaiser and Closs. Van Valkenburg estimates that of the echoes which were measured 80% had a maximum polarization ratio less than 4·0.

The present paper is the result of another attempt at an experimental study of the scattering of radio waves by meteor trails, using linearly polarized waves at a wavelength of 5·4 m, carried out mainly during the summer of 1954.

§ 2. APPARATUS

The same aerial is used for radiating pulses and for receiving echoes. It consists of two perpendicularly polarized four-director Yagi aerials on a common axis, which can be directed to any azimuth and elevation, giving horizontal and vertical planes of polarization for an aerial elevation of 0°. Observations are made during meteor showers when most of the meteors observed are travelling in parallel paths. It can be seen that, when the aerial is pointed in a direction differing from that of the shower radiant by 180° in azimuth and 90° in elevation, the vertically and horizontally polarized aerials give rise respectively to longitudinally and transversely polarized waves at the meteor trails. During observations the aerial is moved in steps of 10° in elevation and at half hourly intervals in azimuth. The errors this practice will introduce into the measurements are considered in § 4.2.

The Yagis each have a measured free-space beam width of 43° between half-power points, where the difference between the power gains is 0·5 dB (figure 1). The cross-coupling between the two aerials is less than -20 dB. The polar diagram in the vertical plane is modified considerably by ground reflections, as indicated in figure 2, which shows the calculated polar diagram for an elevation of the aerial axis of 30°. Reasonable values of the dielectric constant of the soil were assumed ($\kappa = 25 + 19i$ e.s.u.). A common transmit–receive (TR) system is used to connect the aerials to the transmitter and receiver. Because

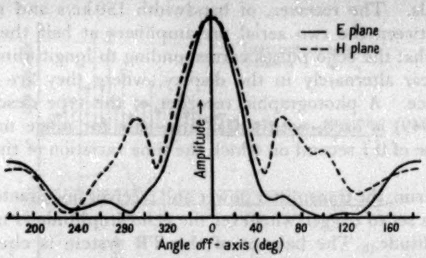


Figure 1. Measured directional diagram of the aerial system (amplitude of electric field against azimuth). In the absence of ground reflections, the E-plane and H-plane polar diagrams correspond in the echo plane to transverse and longitudinal polarization respectively.

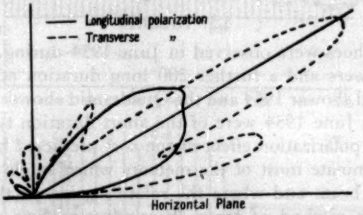


Figure 2. Aerial polar diagram in elevation showing the effect of ground reflections. This is calculated for an aerial elevation of 30°.

the transmitter power is divided equally between the two aerials, while the receiver is switched alternately from one to the other, it is essential that the TR system should be symmetrical as seen from the transmitter while giving negligible cross-coupling between two halves of the system on reception. The transmission line circuit used is shown in figure 3; it has a measured cross-coupling on reception of less than −17 dB.

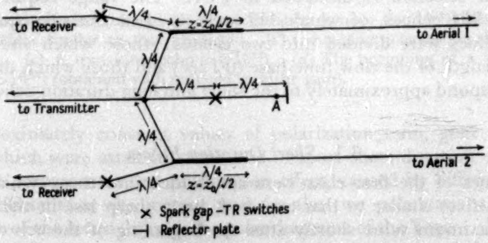


Figure 3. Transmission line system (common TR).

A frequency of 55·3 Mc/s is used (wavelength = 5·4 m), with a pulse length of 8 μsec and a pulse repetition frequency of 600 per second, every fourth pulse being doubled to facilitate range measurement. The transmitter has a peak power of 50 kw, which is divided equally between the two perpendicularly polarized aerials. The receiver, of bandwidth 150 kc/s and noise factor 6 dB, is switched between the two aerial pre-amplifiers at half the pulse repetition frequency, so that the echo pulses corresponding to longitudinal and transverse scattering appear alternately in the display, where they are identified by an automatic device. A photographic recorder of the type described by Davies and Ellyett (1949) is used, with a fast time-base for range measurement, and a slow time-base of 0·1 second on which the time variation of the echo amplitude can be studied.

Before each run, the transmitter power and receiver noise factors are measured. The recorder is set to trigger whenever the echo amplitude is more than double the noise amplitude. The balance of the TR system is checked by feeding a train of radio-frequency pulses into a point A of the transmission line circuit (cf. figure 3), and adjusting the pre-amplifier gains until the two signal levels at the receiver output are equal and are unaltered by interchange of the pre-amplifier inputs.

§ 3. RESULTS

1500 meteor echoes were observed in June 1954 during the daytime Arietid and ζ-Perseid showers and a further 200 long duration echoes were observed during the Geminid shower 1954 and the Quadrantid shower 1955. Many of the echoes observed in June 1954 were of the short duration type and of these the majority showed a polarization effect of the sort predicted by Kaiser and Closs.

In order to eliminate most of the meteors which might be seen in the side lobes of the aerial beam and where the relative gains on the two polarizations are very unequal a method of range discrimination was adopted, only those meteors being accepted which lay within the range limits corresponding to an overall angle of acceptance of 40°. Of these only those echoes were considered for a general analysis which did not have an echo amplitude large enough to give an off-scale deflection on the display, for which the amplitude on both polarizations was greater than 3 × noise for most of the echo, which had a total duration greater than 30 msec, and for which the permanent polarization ratio (see below) was less than 3·0. Only for echoes satisfying all these requirements was it possible to make reliable measurements. The question of bias due to observational selection is discussed in §4.1. The range requirements were satisfied by 924 echoes, of which 112 were used in the subsequent analysis. The 112 echoes were divided into two classes: those which showed a decay during the length of the slow time-base (0·1 sec) and those which did not. The classes correspond approximately to the short and long duration echoes discussed in §1.

3.1. *Short Duration Echoes*

The echoes of the first class were subdivided into those which showed a polarization effect similar to that expected, i.e. a sharp rise in the polarization ratio to a maximum value shortly after the beginning of the echo followed by a decay to a constant value (the theory actually predicts a decay to unity, the fact that the final value was generally greater being accounted for below), and those which showed an anomalous polarization effect in which the polarization ratio increased towards the end of the echo. Two typical short duration echoes, plotted from the records, are shown in figure 4.

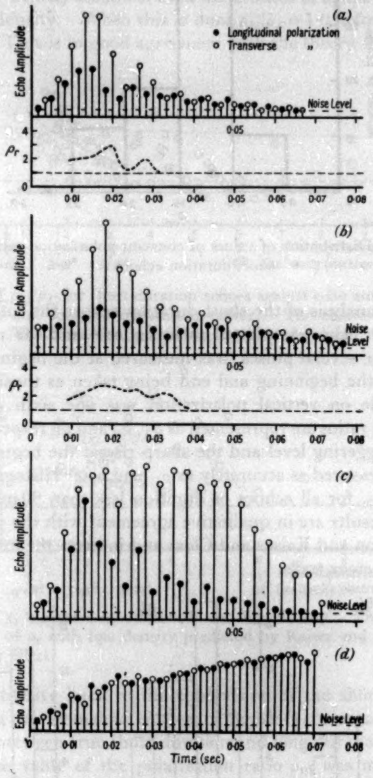


Figure 4. Typical echoes. (*a*) and (*b*) short duration echoes ($\alpha < 10^{12}$) showing a polarization effect similar to that predicted. The changing polarization ratio is plotted for both. (*c*) A transition echo ($\alpha \simeq 10^{12}$). (*d*) The beginning of a long duration echo ($\alpha > 10^{12}$) (obtained with reduced receiver gain).

The approximately constant values of polarization ratio, generally greater than unity, which were attained at the end of most short duration echoes were thought not to represent a real polarization effect, caused by differences in the reflection coefficients at the meteor trail for the two polarizations, but to be a consequence of the inequalities in the polar diagrams of the Yagis. A histogram of the values of these constant polarization ratios for the measured short duration echoes is given in figure 5. The range of values is in fair agreement with that to be expected from the calculated power gain ratios of the Yagis (see figure 2).

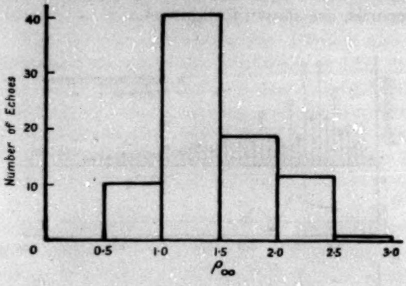


Figure 5. Distribution of values of constant polarization ratio (ρ_∞) for short duration echoes.

In the general analysis of the short duration echoes the following procedure was adopted: each echo was plotted as in figure 4 and the mean polarization ratio, averaged over several pulses, was measured at the beginning, middle and end of each echo, the beginning and end being taken as those points at which the echo amplitude on vertical polarization was one sixth of the maximum amplitude. These ratios are represented as ρ_A, ρ_B and ρ_C respectively. Because of the irregular triggering level and the sharp rise at the beginning of an echo, ρ_A could not be measured as accurately as ρ_B and ρ_C. Histograms of the ratios ρ_B/ρ_A, ρ_B/ρ_C, ρ_A/ρ_C, for all echoes of duration less than 90 msec are shown in figure 6. These results are in qualitative agreement with the predictions of the theories of Herlofson and Kaiser and Closs and indicate the existence of plasma resonance in the meteor trails.

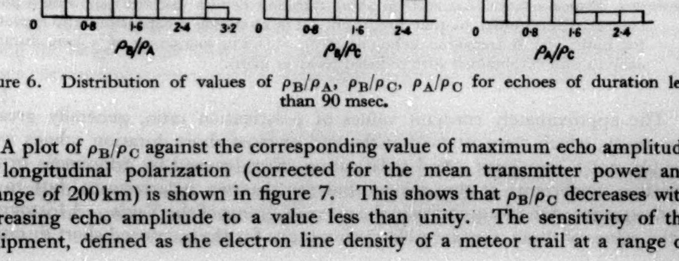


Figure 6. Distribution of values of ρ_B/ρ_A, ρ_B/ρ_C, ρ_A/ρ_C for echoes of duration less than 90 msec.

A plot of ρ_B/ρ_C against the corresponding value of maximum echo amplitude on longitudinal polarization (corrected for the mean transmitter power and a range of 200 km) is shown in figure 7. This shows that ρ_B/ρ_C decreases with increasing echo amplitude to a value less than unity. The sensitivity of the equipment, defined as the electron line density of a meteor trail at a range of

200 km which will produce an echo amplitude of 3 × noise, has been found experimentally to be $\alpha = 10^{11}\,\mathrm{cm}^{-1}$, this being in agreement with an independent estimate from the equipment parameters. If this value of line density is attributed to amplitudes of 3 × noise in figure 7 and the direct proportionality between amplitude and line density assumed, then the abscissa of figure 7 may be replaced by a scale of line density. When this is done $\rho_B/\rho_C = 1{\cdot}0$ is found to correspond to $\alpha = 10^{12}\,\mathrm{cm}^{-1}$. This is in good agreement with the theory of Kaiser and Closs

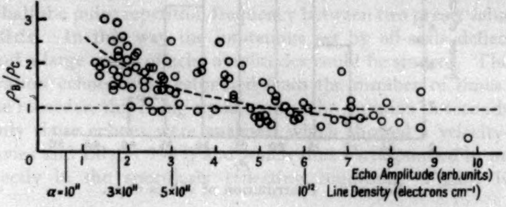


Figure 7. Values of ρ_B/ρ_C for short duration echoes against echo amplitude (corrected to 200 km range and constant transmitter power). The abscissa is given a scale of line density. The chain line is fitted to the plotted values.

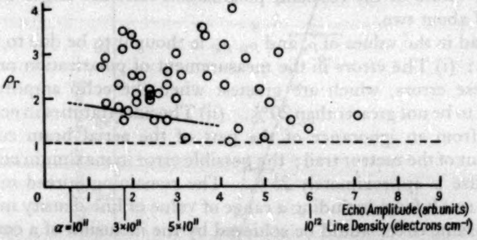


Figure 8. Values of ρ_r against echo amplitude and line density. The full line represents the variation of ρ_r with line density predicted by Kaiser and Closs, neglecting the finite velocity effect.

and gives a quantitative basis to the subdivision of the short duration echoes. It can be seen that the anomalous echoes (cf. figure 4) correspond to line densities in the transition between true short duration and long duration echoes.

The maximum value of the polarization ratio ρ_m was measured for those echoes which showed the expected plasma resonance effect and which had a duration, measured from one-sixth amplitude points, of less than 90 msec. In addition only those echoes were accepted for which the polarization ratio ρ_∞ at the end of the echo had a constant ratio, which was reached before the echo amplitude on either aerial had fallen below about 3 × noise. For the 52 echoes which were used the effect of the inequalities in the two polar diagrams was removed by dividing ρ_m by ρ_∞ for each echo giving ρ_r, the true maximum polarization ratio at resonance. A plot of ρ_r against echo amplitude is given in figure 8. The values of ρ_r show no apparent decrease with increasing amplitude (line density); this is thought to be a consequence of the finite velocity of the meteor (see § 4). A histogram of ρ_r for the measured echoes is given in figure 9.

The theory of Kaiser and Closs predicts a maximum value of the resonant polarization ratio of $\rho_r = 2$, corresponding to $\alpha = 10^{10}\,\text{cm}^{-1}$, with ρ_r decreasing to unity as the line density increases to $10^{12}\,\text{cm}^{-1}$. The minimum line density for which ρ_r was measured was approximately $2 \times 10^{11}\,\text{cm}^{-1}$, for which the theoretical value of ρ_r is 1·7. From figures 8 and 9 it would seem that the experimental values of the resonant polarization ratio are larger than predicted by a factor of about two.

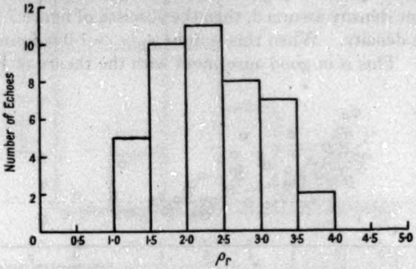


Figure 9. Distribution of values of ρ_r.

The spread in the values of ρ_r and ρ_B/ρ_C is thought to be due to the following uncertainties: (i) The errors in the measurement of polarization ratios from the records; these errors, which are greatest when the echo amplitude is small, are estimated to be not greater than 20%. (ii) The uncertainties in echo amplitude which arise from an ignorance of the part of the aerial beam containing the reflection point of the meteor trail; the possible error in maximum echo amplitude from this cause is approximately 25%. The remaining spread in values of ρ_r is thought to result from including a range of value of line density in the analysis, although the same effect would be achieved by the inclusion of a certain number of sporadic meteors. But since the aerial is always pointing in the most favourable direction for shower meteors and because of the range limitations imposed this number is not very large, the proportion of sporadic meteors being estimated as 15%. The final source of error, which arises from the discontinuous movement of the aerial (causing a misalignment of the Yagis with respect to the meteor trails), is considered in § 4.2.

3.2. *Long Duration Echoes*

Of the 112 echoes which were selected only five showed no decay during the length of the slow time-base (0·1 sec). This is not surprising, because most long duration echoes observed either originate at trails which have been distorted into a specularly reflecting condition by turbulence, and so lie outside the limits of acceptance in range, or else have amplitudes greater than full-scale deflection on the display. More long duration echoes were obtained during the 1955 Quadrantid meteor shower, when observations were made with decreased equipment sensitivity. No long duration echoes were observed which showed the kind of polarization effect predicted for short duration echoes. Most long duration echoes had a constant polarization ratio over their duration, though some showed a changing polarization ratio which accompanied the amplitude fluctuations caused by the atmospheric turbulence later in the lifetime of the echo. A representative long duration echo is given in figure 4: the characteristic slow rise in echo amplitude will be noted, together with the absence of a systematic change in the polarization ratio, with the exceptions noted above.

During the Geminid meteor shower of December 1954, observations were carried out using only the Yagi corresponding to longitudinal polarization, and with the same geometry as before. At the same time, the receiver gain was switched at half the pulse repetition frequency between two preset values differing by about 10 dB. In this way the limitations set by off-scale deflections were overcome and a large range of echo amplitudes could be studied. The durations of long duration echoes were estimated from the number of times each echo triggered the recorder, this being equal to the echo duration in seconds, to within 1·0 sec. Only those echoes were analysed which showed a velocity-diffraction pattern (Davies and Ellyett 1949) and which thus corresponded to meteor trails formed directly in the specularly reflecting condition. Figure 10 shows a logarithmic plot of mean echo amplitude against duration in one second intervals for the 105 echoes of duration greater than one second which were accepted for analysis. It can be seen that the graph is fitted by a relation of the form $A \propto T^{0\cdot31}$ where A is the echo amplitude and T the echo duration. This agrees with an earlier estimate by Greenhow (1952 b) who measured echo durations and amplitudes at a frequency of 36 Mc/s (8·2 m). His measurements lead to the relation $A \propto T^{0\cdot28}$. The significance of these results is discussed in the next section.

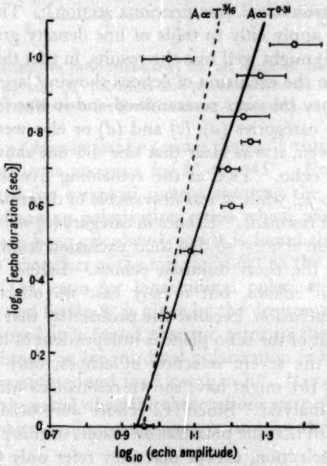


Figure 10. Logarithmic plot of echo amplitude (corrected to 200 km range) in arbitrary units against echo duration. Each point represents a mean for several echoes, the standard deviation being given by the horizontal bar. The full line represents the 'best' (centroid) fit to the points, while the broken line represents the expected fit.

§ 4. Discussion

4.1. *Observational Selection*

As was stated in the previous section, only 12% of the echoes satisfying the range requirements were used in the final analysis. In such circumstances the possibility that the results have been influenced by the selection of data must be carefully considered. The 812 unused echoes were initially rejected for one of the following reasons: (*a*) the longitudinally polarized echo was too weak to measure, (*b*) the transversely polarized echo gave off-scale deflections at some time during the lifetime of the echo, (*c*) both transversely and longitudinally polarized echoes produced off-scale deflections during the lifetime of the echo, (*d*) the echo showed evidence of having been produced after distortion by atmospheric turbulence of a meteor trail not initially formed in the specularly reflecting condition.

The numbers of rejected echoes in each category were: (*a*) 33%, (*b*) 20%, (*c*) 15%, (*d*) 33%. Echoes falling into category (*a*) were caused either by meteors appearing in the side-lobes of the aerial, or by meteor trails of line density less than about 10^{11} electrons cm^{-1} (see previous section). Thus the results in the previous section will apply only to trails of line density greater than this value. Echoes in category (*b*) might well bias the results, in that their rejection from the analysis might lead to the exclusion of echoes showing large polarization effects. The echoes in category (*b*) were re-examined and it was found that only seven did not fall also into categories (*a*), (*c*) and (*d*) or else were not short duration echoes. Of these seven, it was clear that one did not show resonance and that one was a transition echo. Two of the remaining five were resonant echoes, of indeterminate ratio ρ_r, while it was impossible to determine whether the other three were or were not resonant. Echoes in category (*c*) were caused by transition echoes or long-duration echoes, so that their exclusion from the analysis does not affect the results for the short duration echoes. Echoes in category (*d*) were usually long duration echoes, but in any case no observational selection is introduced by their exclusion, because the probability that a section of a meteor trail will be formed out of the echo plane is independent of its line density.

Thus in spite of the severe selection of echoes, only five of the excluded echoes, all in category (*b*), might have shown resonances which should have been included in the final analysis. Since 112 echoes were used in the final analysis, it can safely be assumed that the polarization results in this paper are independent of the observational selection, except that they refer only to a limited range of electron line densities, between 10^{11} and 2×10^{12} electrons cm^{-1}.

4.2. *Misalignment of the Aerial*

The assumption that the meteor trail was parallel to the plane of polarization of the appropriate Yagi was not in general justified as the aerial was moved at half-hourly intervals and was pointed approximately midway between the Arietid and ζ-Perseid radiants.

It can be shown that if the plane of polarization of the longitudinal Yagi is inclined at an angle θ to the direction of the trail then the polarization ratio ρ_θ' which will be measured is given by

$$\rho_\theta' = \frac{(\sin^2\theta + \sin\theta\cos\theta) + \rho_0' \exp(i\psi_0')(\cos^2\theta - \sin\theta\cos\theta)}{(\cos^2\theta + \sin\theta\cos\theta) + \rho_0' \exp(i\psi_0')(\sin^2\theta - \sin\theta\cos\theta)} \qquad \ldots\ldots(2)$$

where ρ_0' is the polarization ratio and ψ_0' the phase difference between the signals which would be measured on transverse and longitudinal polarization if $\theta = 0°$; ρ_0' and ψ_0' differ from ρ and ψ because of the finite velocity effect. A plot of ρ_θ' against ρ_0' for $\theta = 20°$ and for several values of ψ_0' is given in figure 11. For small ψ_0' and large ρ_0', ρ_θ' may in fact be greater than ρ_0'; this is because the second term in the denominator is negative for $0° < \theta < 45°$. From the values of ρ_0' and ψ_0' to be expected for a typical meteor showing the finite velocity effects (see figure 13), the changing polarization ratios which would be measured for $\theta = 0°$, 20° and 30° have been calculated, and it is found that for values of θ not greater than about 20° the effect of the misalignment on the observed polarization ratio is negligible. The ratio for longitudinal polarization of the amplitude observed to that expected for $\theta = 0°$ is given by the denominator of expression (2). If typical values are used it is found that the error in the measurement of the maximum echo amplitude on longitudinal polarization is less than 10% for the polarization ratios encountered and for θ less than 30°. It is not expected therefore that misalignment of the aerial when the observations were obtained will introduce appreciable errors in the measurement of maximum echo amplitude or polarization ratio.

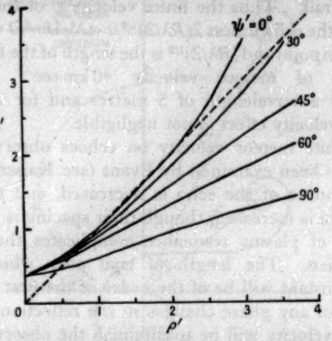


Figure 11. ρ_θ' the polarization ratio measured for an angle of misalignment θ for $\theta = 20°$ as a function of ρ_0' and ψ_0' the ratio and phase difference respectively between the overall reflection coefficients for $\theta = 0°$, taking into account the finite velocity effect.

An effect similar to that produced by misalignment would be produced by rotation of the plane of polarization of the wave on its passage through the D and lower E regions of the ionosphere. An estimate shows that for the directions in which the aerial was pointed during the observations discussed in this paper the rotation of the plane of polarization will always be less than 10° for meteors below 110 km in height. Accordingly the effect has been neglected in this analysis. It should be noted that for aerial directions close to that of the terrestrial magnetic field, and for frequencies below about 70 Mc/s, the rotation can be of the order of one radian during the daytime, when the D regions have greatest ionization.

4.3. *Finite Velocity Effect*

When the time for the meteor to cross the first few Fresnel zones on either side of the specular reflection point is not small compared with the lifetime of the echo, it is incorrect to assume that at any instant the same reflection coefficients characterize every part of the ionized trail. In particular, the variation with time of the amplitude and phase of the echo will not necessarily be similar to the variation with time of the amplitude and phase of the reflection coefficient of any portion of the trail. Thus the finite velocity v of the meteor will have an effect on the form of the echo unless $2(R\lambda/2)^{1/2}/v \ll \lambda^2/16\pi^2 D$ where R is the range of the specular reflection point and $(R\lambda/2)^{1/2}$ is the length of the first Fresnel zone. For the typical values of meteor velocity 40 km sec^{-1}, diffusion coefficient 3×10^4 cm^2 sec^{-1}, for a wavelength of 5 metres and for $R = 200$ km, it can be seen that the finite velocity effect is not negligible.

The effect of finite meteor velocity on echoes observed with longitudinal polarization only has been examined by Evans (see Kaiser 1953) who finds that the maximum amplitude of the echo is decreased, and that the depth of the Fresnel fringe pattern is increased, though their spacing is unaltered.

The occurrence of plasma resonance complicates the effects observed on transverse polarization. The length of trail over which plasma resonance is occurring at any instant will be of the order of the first Fresnel zone only, so that in the absence of any phase changes in the reflection coefficient, the effect of the finite meteor velocity will be to diminish the observed polarization ratio. When the phase changes accompanying the resonance are taken into account, however, it is found that in some cases the finite velocity effect may actually increase the observed polarization ratio.

The echo amplitudes expected from a meteor trail are given by

$$E_L = \text{constant} \times \int_{-\infty}^{n_0} \exp\{-k\delta(n_0 - n)\} \exp(i\phi)\, dn \qquad \text{(longitudinal)}$$

and

$$E_T = \text{constant} \times \int_{-\infty}^{n_0} \rho(n) \exp\{-k\delta(n_0 - n)\} \exp i(\phi + \psi)\, dn \qquad \text{(transverse)}$$

where n is the distance in units of $(R\lambda)^{1/2}$ from the specular reflection point, n_0 is the corresponding distance to the instantaneous head of the trail, v is the meteor velocity, $k = 16\pi^2 D/\lambda^2$, $\delta = (R\lambda)^{1/2}/v$, $\phi(=2\pi n^2)$ is the phase angle of the element dn with respect to the phase at $n = 0$, ψ is the phase lead of the transverse reflection coefficient and ρ is the ratio of the transverse to the longitudinal reflection coefficients at the element dn. The integrals can be evaluated numerically for a given value of n_0, and a phase amplitude diagram, similar to the well-known Cornu spiral, can be plotted. Although Kaiser and Closs did not calculate the way in which ψ varies with time, a sufficiently good approximation for the present purpose is given by assuming that ψ decays exponentially from $+180°$ to $0°$, passing through the value 45° when ρ passes through its maximum. Figure 12 shows a phase–amplitude diagram for the values $n_0 = 1{\cdot}0$, $v = 40$ km sec^{-1}, $D = 3\times10^4$ cm^2 sec^{-1}, and $\alpha = 3\times10^{11}$ electron cm^{-1} ($\rho_r = 1{\cdot}7$ is predicted for this line density). It can be seen that the variation of ψ along the trail tends to 'unwind' the vector spiral for positive values of n, so increasing the echo amplitude. If the amplitude and phases of E_L and E_T evaluated from plots such as this for different values of n_0 are themselves displayed on a phase–amplitude diagram, a plot is obtained which shows the variation with time of the amplitude and phase of the echo (figure 13 (a)). From this the variation of amplitude alone can be obtained (figure 13 (b)).

From figure 13 (*b*), it would seem that the overall effect of the finite meteor velocity is to increase the observed maximum polarization ratio and to alter the velocity diffraction pattern observed with transverse polarization. It is thus fairly certain that the finite velocity effect is responsible for the observed polarization ratios being on the whole greater than those expected from the simple application of Kaiser and Closs's theory in which it is assumed that the whole meteor trail is formed instantaneously.

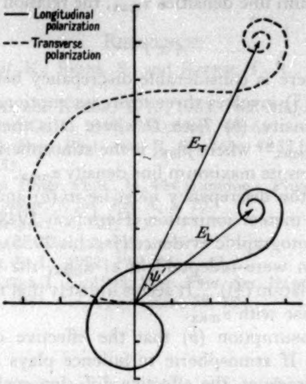


Figure 12. Phase–amplitude vector diagram plotted for the typical echo ($n_0=1{\cdot}0$).

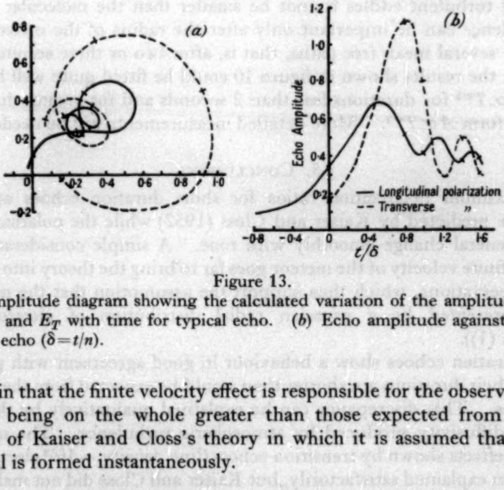


Figure 13.

(*a*) Phase–amplitude diagram showing the calculated variation of the amplitude and phase of E_L and E_T with time for typical echo. (*b*) Echo amplitude against time for the same echo ($\delta=t/n$).

4.4. *The Amplitude–Duration Relationship for Long Duration Echoes*

It has been shown (cf. figure 10) that the amplitude A and duration T of long duration echoes are related empirically by the expression

$$A \propto T^{0{\cdot}3}. \qquad \ldots\ldots (3)$$

For a constant diffusion coefficient, the predicted relation is

$$A \propto T^{1/4} \qquad \ldots\ldots(4)$$

but because the point of maximum ionization should occur lower in the atmosphere with increasing maximum line densities $\alpha_{\max}$, the relation (4) is modified (Kaiser 1953):

$$A \propto T^{3/16}. \qquad \ldots\ldots(5)$$

It can be seen that there is considerable discrepancy between observation and prediction. Relation (5) involves three main assumptions: (*a*) $A \propto \alpha^{1/4}$ where α is the electron line density, (*b*) $T \propto \alpha/D$ where D is the atmospheric diffusion coefficient, (*c*) $p_{\max} \propto \alpha_{\max}{}^{1/3}$ where $p_{\max}$ is the atmospheric pressure at the point where a meteor produces its maximum line density $\alpha_{\max}$. Assumption (*a*) seems well founded, so that the discrepancy must lie in (*b*) and (*c*). Assumption (*c*) rests on the theory of meteor ionization (Herlofson 1948) on which doubt has been cast by recent photographic evidence (Jacchia 1955), but even if the height of maximum ionization were independent of $\alpha_{\max}$, the exponent of T should be no greater than $\frac{1}{4}$ (relation (4)). It seems unlikely that the height of maximum ionization would increase with $\alpha_{\max}$.

It is implied in assumption (*b*) that the effective diffusion coefficient is independent of time. If atmospheric turbulence plays any part in spreading the initial ionization, however, the effective diffusion coefficient would increase with time. Greenhow (1952 a) has already suggested that turbulence plays a part in shortening echo durations greater than about 10 seconds. But since the smallest turbulent eddies cannot be smaller than the molecular mean free path, turbulence can be important only after the radius of the meteor trail has increased to several mean free paths, that is, after two or three seconds. It can be seen that the results shown in figure 10 could be fitted quite well by a law of the form $A \propto T^{0 \cdot 2}$ for durations less than 2 seconds and for greater durations by a law of the form $A \propto T^{0 \cdot 3}$. More detailed measurements will be needed to settle this point.

§ 5. CONCLUSIONS

The maximum polarization ratios for short duration echoes are roughly double those predicted by Kaiser and Closs (1952) while the polarization ratios do not in general change smoothly with time. A simple consideration of the effect of the finite velocity of the meteor goes far to bring the theory into agreement with the observations, which thus support the assumption that the meteor trail can be represented by a gaussian radial distribution of electron density (cf. equation (1)).

Long duration echoes show a behaviour in good agreement with prediction, except that their durations are shorter than would be expected from the amplitude of the echoes. This discrepancy can be explained qualitatively by the increase in effective diffusivity produced by atmospheric turbulence. The unexpected polarization effects shown by transition echoes (line density $\sim 10^{12}$ electrons cm^{-1}) have not been explained satisfactorily, but Kaiser and Closs did not make detailed calculations for this range of line densities. It should be noted that when meteor echoes are studied without distinction between the states of polarization, errors may be introduced by resonance effects into (i) the velocity measurements from the Fresnel diffraction pattern, (ii) the measurement of decay time of short duration echoes.

ACKNOWLEDGMENTS

We wish to thank in particular Dr. T. R. Kaiser for his constant advice and help. We are grateful also to Dr. J. S. Greenhow for many helpful discussions, and to Professor A. C. B. Lovell for facilities at the Jodrell Bank Experimental Station.

REFERENCES

BROWNE, I. C., BULLOUGH, K., EVANS, S., and KAISER, T. R., 1956, *Proc. Phys. Soc.* B, **69,** 83.

CLOSS, R. L., CLEGG, J. A., and KAISER, T. R., 1953, *Phil. Mag.*, **44,** 313.

DAVIES, J. G., and ELLYETT, C. D., 1949, *Phil. Mag.*, **40,** 614.

GREENHOW, J. S., 1952 a, *Proc. Phys. Soc.* B, **65,** 169; 1952 b, *Dissertation*, University of Manchester, p. 55.

HERLOFSON, N., 1948, *Rep. Progr. Phys.*, **11,** 444 (London : Physical Society).

HUXLEY, L. G. H., 1952, *Austr. J. Sci. Res.* A, **5,** 10.

JACCHIA, L. G., 1955, *Astrophys. J.*, **121,** 521.

KAISER, T. R., 1953, *Advanc. Phys.*, **2,** 495.

KAISER, T. R., and CLOSS, R. L., 1952, *Phil. Mag.*, **43,** 1.

LOVELL, A. C. B., and CLEGG, J. A., 1948, *Proc. Phys. Soc.* B, **60,** 491.

VAN VALKENBURG, M. E., 1954, *J. Geophys. Res.*, **59,** 359.

RESEARCH NOTES

Ion Oscillations in a Cathode Potential Minimum

By K. G. Emeléus and N. R. Daly
Queen's University, Belfast

MS. received 13th July 1955, and in amended form 25th October 1955

Discharges from hot cathodes through gas at low pressure may be sources of noise in the kc/s to Mc/s frequency band, with superposed oscillations possessing definite frequencies (Cobine and Gallagher 1947, Martin and Woods 1951). When the current is not saturated, it has been found repeatedly in this laboratory that the frequency ν of the regular oscillations is nearly proportional to the square root of the current $|I|$. It is difficult to trace with certainty the origin of an oscillation in a complex system of space charges and plasma, but it is perhaps interesting to note that a relation of this type might be expected for the small amplitude oscillation of ions in the potential minimum in front of the cathode, when the current is small compared with the saturation current.

In the plane case, we will assume that in the region of the plane, which we will call $x=0$, where the potential V has a minimum, a unit volume contains n_0 electrons each with charge $-\mathbf{e}$, and relatively few ions. Then

$$\left(\frac{d^2V}{dx^2}\right)_{x=0}=4\pi n_0\mathbf{e}. \qquad \ldots\ldots (1)$$

If T is the temperature of the cathode and electrons, the current i flowing between cathode and anode per unit area perpendicular to x will be, under the conditions stated at the end of the first paragraph (cf. Weizel 1950), nearly

$$i=-n_0\mathbf{e}(\boldsymbol{k}T/2\pi\mathbf{m})^{1/2}. \qquad \ldots\ldots (2)$$

Integrating equation (1) and substituting for n_0 from (2) gives

$$\frac{dV}{dx}=-4\pi ix(2\pi\mathbf{m}/\boldsymbol{k}T)^{1/2}. \qquad \ldots\ldots (3)$$

It follows that the frequency of small amplitude oscillations of singly charged positive ions of mass M about the minimum will be

$$\nu=\left\{\frac{\mathbf{e}(-i)}{\pi M}\left(\frac{2\pi\mathbf{m}}{\boldsymbol{k}T}\right)^{1/2}\right\}^{1/2}, \qquad \ldots\ldots (4)$$

showing that for a constant cathode temperature and ion mass, ν is proportional to $(-i)^{1/2}$ and hence to $|I|^{1/2}$.

Writing equation (4) in the alternative form

$$\nu=\{n_0\mathbf{e}^2/\pi M\}^{1/2}, \qquad \ldots\ldots (5)$$

it is seen that ν found in this way is identical with the upper limiting frequency of oscillation of positive ions in a plasma with concentrations n_0 of ions and electrons.

The analysis for a cylindrical cathode will be less simple, on account of the variation of the area of the minimum potential surface with current. In many cases this is likely to be small (Wheatcroft 1940), and a relation of the form (4) still approximately true. If the amplitude of the oscillations and the contribution of the positive ions to the space charge become large, White's experiments (1953) show that more complicated conditions result.

REFERENCES

COBINE, J. D., and GALLAGHER, C. J., 1947, *J. Appl. Phys.*, **18**, 110.
MARTIN, H., and WOODS, H. A., 1951, *Proc. Phys. Soc.* B, **65**, 281.
WEIZEL, W., 1950, *Lehrbuch der theoretischen Physik*, Vol. II (Berlin : Springer), p. 1287.
WHEATCROFT, E. L. E., 1940, *J. Instn Elect. Engrs*, **86**, 473.
WHITE, O., 1953, *Proc. Phys. Soc.* B, **66**, 577.

Photon-radiative Recombination in PbSe, PbTe and PbS

BY I. M. MACKINTOSH
Department of Physics, University of Nottingham

Communicated by L. F. Bates; MS. received 6th June 1955 and in final form 29th August 1955

THE increasing availability of pure single crystals of semiconductors of relatively small energy gap, such as InSb, PbS etc., has led to a more exact knowledge of the fundamental properties of these materials, and has made possible a closer investigation of their subsidiary properties. In particular, there is experimental evidence that in all the semiconductors of small energy gap so far investigated, the recombination lifetime of free electrons and holes is very short (of the order microseconds and less). One of the recombination mechanisms which may be of importance in such materials is direct electron–hole recombination resulting in the emission of a single photon, and it is the purpose of this note to obtain an estimate of the lifetime due to this process in the semiconductors of the PbS group.

van Roosbroeck and Shockley (1954) have treated photon-radiative recombination in germanium by applying the principle of detailed balance and equating absorption and generation rates at thermal equilibrium. They show that the total rate R of radiative recombination of electrons and holes per unit volume at thermal equilibrium and at 300°K, may be expressed in the form

$$R = 1{\cdot}78 \times 10^{22} \int_0^\infty n^3 K \frac{u^3}{e^u - 1}\, du \text{ cm}^{-3}\text{ sec}^{-1} \quad \ldots\ldots (1)$$

where n is the refractive index, K is the absorption index, and $u = h\nu/kT$.

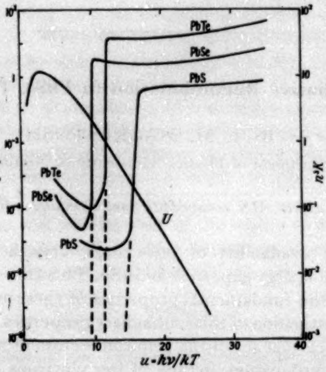
Evaluation of R thus requires numerical integration of the product of n^3K and the dimensionless quantity $U = u^3/(e^u - 1)$ which is the same for all materials. The lower limit of integration may be taken as the value of u which corresponds to the long-wavelength absorption limit, since photons of longer wavelength are absorbed by other mechanisms, such as free-carrier absorption.

The relation between the lifetime and the recombination rate has been obtained for the non-degenerate case by van Roosbroeck (1953), and for the degenerate case by Mackintosh and Allen (1955). The intrinsic photon-radiative lifetime τ_i is given by

$$\tau_i = n_i/2R \qquad \text{......(2)}$$

where n_i is the concentration of intrinsic electrons.

The figure shows the variation of the factor U with u, and the variation of n^3K with u for the compounds PbSe, PbTe and PbS at 300°K. Following experience gained in the analysis of InSb, a direct extrapolation of the 'edge' has been made, without attempting to compute the effect of long-wavelength absorption. The values of n^3K were calculated from the published absorption data of Avery (1954) and Gibson (1952). Avery uses reflection techniques to determine the absorption coefficient α in the short-wavelength range where the value is high (10^4 cm), and claims a resolution of $\pm 0{\cdot}1\,\mu$; Gibson determines the absorption coefficient on the long-wavelength side of the absorption edge by transmission techniques and claims a resolution of $\pm 0{\cdot}2\,\mu$. In order to bring the two sets of measurements into approximate agreement at the absorption edge, it was found necessary to shift one or both curves along the wavelength axis by amounts not exceeding the resolutions claimed. This procedure is much the same as that used by Avery in building up a composite curve of absorption coefficient against photon energy for PbS. Gibson's values were shifted $0{\cdot}1\,\mu$ for PbS, and $0{\cdot}2\,\mu$ for PbTe and PbSe, all towards the longer wavelengths; Avery's values were shifted $0{\cdot}1\,\mu$ for PbTe towards the shorter wavelengths.



Dependence of n^3K on $u = h\nu/kT$ for PbSe, PbTe and PbS at 300°K.

The values of R found by subsequent numerical integration are shown in the table. No effect on the position of the absorption edge of varying impurity concentrations has been observed in these compounds, and it may therefore be assumed that these values of R represent the photon-radiative recombination rate in intrinsic material.

Calculation of n_i at 300°K followed the lines suggested by Smith (1954). If E_G is the energy gap at T°K, m_n and m_p are the effective masses of electrons and holes respectively, and m is the mass of the free electron, then we have

$$n_i = 4{\cdot}82 \times 10^{15} \times T^{3/2} \left(\frac{m_n m_p}{m^2}\right)^{3/4} \exp\left(-E_G/2\boldsymbol{k}T\right). \qquad \ldots\ldots(3)$$

This expression is given by Shockley (1950, p. 245) and is accurate only for non-degenerate samples. The factor $(m_n\, m_p/m^2)^{1/2}$ has been obtained for each of the compounds using the data on the temperature dependence of the mobility ratio b, and on the dependence of $(m_n\, m_p/m^2)^{1/2}$ on b, provided by Smith (1954, fig. 3 and table III).

The effective masses are also related to the mobility ratio by the expression (Shockley 1950, p. 287),

$$\frac{\mu_n}{\mu_p} = b = \left(\frac{m_p}{m_n}\right)^{5/2} \qquad \ldots\ldots\ldots(4)$$

and combination of (3) and (4) enables values to be obtained for the respective effective masses m_n and m_p. For the PbS group of semiconductors these values will necessarily be tentative due to the large uncertainty in b, and to the probability that the scattering law on which (4) is based does not hold in these materials.

	R (cm^{-3} sec^{-1})	b	$\dfrac{(m_n\, m_p)^{\frac{1}{2}}}{m}$	$\dfrac{m_n}{m}$	$\dfrac{m_p}{m}$	E_G (ev)	n_i (cm^{-3})	τ_i (μsec)
PbSe	$2{\cdot}0 \times 10^{22}$	1·3	0·16	0·15	0·17	0·22	$2{\cdot}3 \times 10^{16}$	0·6
PbTe	$8{\cdot}6 \times 10^{21}$	1·8	0·19	0·17	0·21	0·26	$1{\cdot}3 \times 10^{16}$	0·8
PbS	$4{\cdot}8 \times 10^{19}$	1·5	0·13	0·12	0·14	0·30	$3{\cdot}5 \times 10^{15}$	40

The values of E_G used (see table) are those arrived at by the comprehensive analysis of Smith. Although derived in part from the absorption curves of Gibson, which in this analysis have been shifted along the wavelength axis, it is thought that the values of E_G are sufficiently uncertain in themselves to warrant neglecting a small change on these grounds. If E_G and $(m_n\, m_p/m^2)^{1/2}$ are known, n_i and hence τ_i may be calculated. Values of b, $(m_n\, m_p/m^2)^{1/2}$, m_n, m_p, E_G, n_i, and τ_i, at 300°K are also shown in the table. For comparison, τ_i has been calculated to be 0·75 sec for Ge (van Roosbroeck and Shockley 1954) and $0{\cdot}75 \times 10^{-6}$ sec for InSb (Mackintosh and Allen 1955).

Experimental results are available only for PbS, and Smith states that the longest observed lifetime in this material, using the light-spot method, was 8 μsec. Comparison with the value $\tau_i = 40$ μsec indicates that the assumption that photon-radiative recombination will be an important process in materials of small energy gap (Mackintosh and Allen 1955) is further supported by the results for PbS.

Lifetimes in PbS in the range 10^{-10} sec to 10^{-6} sec have been measured by Moss using the photo-electromagnetic effect, although the interpretation of the photo-electromagnetic effect itself, and of lifetimes less than about 10^{-8} sec, is uncertain (Allen and Mackintosh 1955). Moss gives a curve of lifetime against resistivity which suggests that for PbS the Auger effect is the predominant recombination mechanism. However, taking the values for PbS at 300°K of $\mu_n + \mu_p \sim 80 \times 10^2$ cm^2 volt^{-1} sec^{-1}, and $n_i \sim 3{\cdot}5 \times 10^{15}$ cm^{-3}, a simple calculation gives $\sigma_i \sim 0{\cdot}5$ ohm^{-1} cm^{-1}. Thus a resistivity of the order of 2·0 ohm cm corresponds to an intrinsic photon-radiative lifetime of the order of 40 μsec, which is significantly less than the lifetime obtained by extrapolating Moss's curve for the Auger effect. This seems to support the view expressed by Smith that the Auger effect is unlikely to be the most important process in pure material.

Decay times much in excess of these values of τ_1 have been reported for PbS and PbSe photocells, but in view of the impure and polycrystalline nature of the materials in these cells, it is likely that the decay times represent the effect of trapping processes and give no information on the bulk properties.

The final results for the intrinsic photon-radiative lifetime τ_i depend on the values of R and n_i, in accordance with equation (2). The value of the numerical integral, from which R is obtained, depends both on the position of the absorption edge and on the short-wavelength values of n^3K, for which experimental data are available. Although the latter are probably the more inaccurate due to difficulties in reflection techniques, the fall of the factor U is already predominant in the region where the n^3K curves flatten out, and the result of any errors in these values is likely to be small. The position of the absorption edge, on the other hand, has been investigated by several observers and is known relatively accurately.

The value of n_i depends both on E_G and the factor $(m_n m_p/m^2)$, both of which are somewhat uncertain. However, by assuming an error of ±10% in the values of E_G and taking into account the inaccuracies in $(m_n m_p/m^2)$, the results for n_i are estimated to be accurate to within factors of 2, 3 and 5 for PbSe, PbTe and PbS respectively.

The possibility of any contribution by exciton states to the absorption spectrum has been neglected in view of the evidence (Smith 1954) that the limits of photoconductivity agree reasonably well with the position of the absorption edge in these compounds.

The results of this analysis thus indicate that in the absence of traps, 0·6, 0·8 and 40 μsec represent the maximum attainable lifetimes in the compounds PbSe, PbTe and PbS to within factors of 2, 3 and 5 respectively.

ACKNOWLEDGMENTS

The author is grateful to Professor L. F. Bates for encouragement and for the facilities of his laboratory, to Dr. K. W. H. Stevens for helpful suggestions, to Dr. A. F. Gibson and Dr. D. G. Avery for use of their absorption data, and to the Admiralty for sponsoring this work. Considerable benefit has accrued from discussion with Mr. J. W. Allen, to whom grateful acknowledgment is made.

REFERENCES

ALLEN, J. W., and MACKINTOSH, I. M., 1955, *J. Electronics*, **1**, 138.
AVERY, D. G., 1954, *Proc. Phys. Soc.* B, **67**, 2.
GIBSON, A. F., 1952, *Proc. Phys. Soc.* B, **65**, 378.
MACKINTOSH, I. M., and ALLEN, J. W., 1955, *Proc. Phys. Soc.* B, **68**, 985.
MOSS, T. S., 1953, *Proc. Phys. Soc.* B, **66**, 993.
VAN ROOSBROECK, W., 1953, *Phys. Rev.*, **91**, 282.
VAN ROOSBROECK, W., and SHOCKLEY, W., 1954, *Phys. Rev.*, **94**, 1558.
SHOCKLEY, W., 1950, *Electrons and Holes in Semiconductors* (New York : van Nostrand).
SMITH, R. A., 1954, *Physica*, **20**, 910.

LETTERS TO THE EDITOR

The Distribution of Impurity in a Semi-infinite Solidified Melt

The distribution of impurity in a solid grown from the melt has been treated by several authors (Tiller *et al.* 1953, Pohl 1954).

Tiller *et al.* made use of a simplifying assumption which greatly facilitated the actual calculation. Pohl carried out a straightforward solution of the mathematical problem. He assumed a diffusion region in the melt of infinite length. By means of Laplace transformation of the differential equation mentioned below, Pohl obtained a solution in which the impurity concentration C_s in the solid is given as an infinite series.

It is possible however, as will be shown here, to derive an expression for C_s in closed form based on the same model.

The solute concentration C_l in the liquid is governed by the equation

$$\frac{\partial C_l}{\partial t}=D\frac{\partial^2 C_l}{\partial x^2}+V\frac{\partial C_l}{\partial x} \qquad \ldots\ldots(1)$$

if $x=0$ lies in the solid–liquid interface which moves with a constant velocity V. Equation (1) is readily transformed into

$$\frac{\partial f}{\partial t}=D\frac{\partial^2 f}{\partial x^2} \qquad \ldots\ldots(2)$$

by substituting

$$C_l(x,t)=\text{const.}\times\exp\left\{-\frac{Vx}{2D}-\frac{V^2t}{4D}\right\}f(x,t)+C_0. \qquad \ldots\ldots(3)$$

A particular solution of (2) describing a transient process is

$$f(x,t)=\frac{1}{2(Dt)^{1/2}}\int_{-\infty}^{+\infty}f(x',0)\exp\left\{-\frac{(x-x')^2}{4Dt}\right\}dx'. \qquad \ldots\ldots(4)$$

The zero condition $f(x,0)$ is known for $x\geqslant 0$:

$$C_l(x,0)=C_0 \text{ and therefore } f(x,0)=0.$$

In the region $x<0$ one has to find a function $f(x,0)$ which after integration (4) and substitution in (3) satisfies the boundary condition

$$x=0;\ (1-k)VC_l(0,t)=-D\frac{\partial C_l}{\partial x}. \qquad \ldots\ldots(5)$$

The desired function proves to be:

$$f(x,0)=-\frac{k}{1-k}\exp\left\{\frac{Vx}{2D}\right\}+\exp\left\{-\frac{Vx}{2D}\right\}-\frac{1-2k}{1-k}\exp\left\{-(1-2k)\frac{Vx}{2D}\right\} \qquad (x<0). \qquad \ldots\ldots(6)$$

As now the transient distribution in the liquid is known, substitution of $x=0$ in (3) and multiplication of $C_l(0,t)$ by the equilibrium solubility constant k gives the transient concentration in the solid

$$C_s=kC_l(0,t)=\tfrac{1}{2}C_0[1+\operatorname{erf}(\tau)^{1/2}-(1-2k)\exp[-4k(1-k)\tau]\{1+\operatorname{erf}[(1-2k)\tau^{1/2}]\}]. \qquad \ldots\ldots(7)$$

where τ denotes $V^2t/4D$ and Vt is the positional coordinate in the solid.

Mr. Penning (private communication) of this laboratory calculated the transient distribution C_s by separation of variables in equation (1). A series expression is obtained which may be applied to cases of finite as well as infinite regions of diffusion. There was good agreement of numerical values for the infinite case between Penning's expression and equation (7).

Very recently Hulme (1955) obtained an expression for C_s which is identical to (7), using Laplace transformation in order to solve differential equation (1).

Details of our method of calculation and numerical values of C_s for

$$0{\cdot}01 \leqslant V^2t/4D \leqslant 1000 \quad \text{and} \quad 0{\cdot}001 \leqslant k \leqslant 10$$

will appear in a forthcoming issue of the *Philips Research Reports*.

Philips Research Laboratories,
Eindhoven, Netherlands.
20th October 1955.

O. W. Memelink.

Hulme, K. F., 1955, *Proc. Phys. Soc.* B, **66,** 393.
Pohl, R. G., 1954, *J. Appl. Phys.*, **25,** 1170.
Tiller, W. A., Jackson, K. A., Rutter, J. W., and Chalmers, B., 1953, *Acta Metallurgica*, **1,** 428.

Radiation from Gaseous Arcs

An expression for the radiation loss from the positive column was included in the energy balance equation for high pressure arcs stated by Champion (1953). Elenbaas (private communication) has suggested that the factor n_e (the electron density) might be replaced by n_0 (the concentration of atoms in the ground state) in this relation. From the following it will be seen that both forms of the expression hold, but over different ranges of conditions.

When there is thermal excitation the number of atoms excited to various levels x is

$$\sum n_x = \sum n_0 \frac{g}{g_0} \exp(-\boldsymbol{e}V_x/\boldsymbol{k}T).$$

If R_x is the transition probability from the level x the radiation (assuming no absorption) is

$$\sum n_x R_x = \frac{n_0}{g_0} \sum g R_x \exp(-\boldsymbol{e}V_x/\boldsymbol{k}T) = B(T) n_0 \exp(-\boldsymbol{e}V_{ex}/\boldsymbol{k}T) \quad \ldots\ldots(1)$$

where B can vary appreciably with temperature, and hence, position in an arc. V_{ex} denotes an effective excitation potential, which is a weighted mean of the individual excitation potentials V_x.

When excitation is caused mainly by electron and ion collisions and absorption is negligible the radiation must equal the rate of excitation, which may be expressed in the form

$$\int f(v) n_e n_0 Q_{ex}(v) v\, 4\pi v^2\, dv = n_e n_0 \int_{V_{ex}}^{\infty} \exp(-\boldsymbol{m}v^2/2\boldsymbol{k}T) Q_{ex}(v) 4\pi v^3\, dv$$

$$= A(T) n_e n_0 \exp(-\boldsymbol{e}V_{ex}/\boldsymbol{k}T) \quad \ldots\ldots(2)$$

where $Q_{ex}(v)$ is the total excitation collision cross section, and A varies only slightly with temperature. The variation of n_e with position in the arc is much greater than that of n_0. This expression will be accurate unless excitation by collisions of atoms is appreciable. However, in that case another term (including n_0 but not n_e) can be added to include these excitations.

Expression (1), deduced for thermal excitation, holds when conditions approach those for thermal statistical equilibrium. However, in the limit of complete equilibrium the principle of detailed balancing requires that as much radiation is absorbed as emitted. Thus expression (1) holds over a range of conditions, but does not hold either too near or too far from thermal statistical equilibrium.

Expression (2) should hold over a wide range of conditions when excitation by electron and ion collisions predominate and, when corrected by the additional term to take into account excitation by atom collisions, should cover the whole range of conditions except those where absorption is large, i.e. near complete thermal statistical equilibrium. Expression (2) (without the second term) probably holds at lower pressures than expression (1), i.e. in a region where the electron and ion temperatures are at least slightly above the temperature of the neutral atoms. It should be possible to investigate these radiation terms experimentally.

Department of Physics,
Tufts University,
Medford, Massachusetts.
20th September 1955.

K. S. W. CHAMPION.

CHAMPION, K. S. W., 1953, *Proc. Phys. Soc.* B, **66**, 169.

REVIEWS OF BOOKS

Imagerie géométrique. Aberrations (Traité d'optique instrumentale), by A. MARÉCHAL. Pp. xii + 244. (Paris: Editions de la Revue d'Optique, 1952). 1680 *fr*.

This book constitutes volume 1 of the first part of a projected treatise on optical instruments. The first section, comprising some 60 pages, gives a straightforward account of the general laws of geometrical optics, some simple stigmatic systems, gaussian optics, and the functions of diaphragms in limiting the aperture and field of view. The second, and longer part, takes up the question of aberrations, dealing with each aberration in turn and giving both the ray and wavefront interpretations—except for some reason in the case of chromatic aberration. The discussion of aberrations concludes with a treatment of the Seidel aberration coefficients for axially symmetrical systems and their dependence on the position of the pupil and the magnification, together with the aberrations of a slightly decentred system. A particularly good account of the experimental study of aberrations is given. The final chapters deal with erecting and other prismatic systems, including their aberrations, the geometrical aberration of gratings, non-symmetrical astigmatic systems, and (very briefly) with electron optical systems. Little is said concerning the correction of aberrations, this being reserved for a future volume.

The mathematical treatments are clear and direct, and the many diagrams are a useful adjunct to the text. As a basis for a course on applied optics it leaves little to be desired. H. H. HOPKINS.

Mess- und Prüfmethoden der optischen Fertigung, Vol. 1, by J. PICHT. Pp. x + 175. (Berlin: Akademie Verlag.)

Professor Picht is best known in England as a theoretical physicist and the author of a standard work on the diffraction theory of optical image formation. It is, therefore, surprising to encounter a book of laboratory and workshop practice coming from his pen. We find, nevertheless, a 'professional' manner which clearly derives from both practising and teaching the subject-matter of the book.

The first 73 pages form a comprehensive account of the methods used for the measurement of the optical constants of materials, the accuracy and curvatures of surfaces. The next 23 pages describe the testing of glass for defects such as bubbles and strain, and also photometric measurements on optical systems. The final 72 pages are concerned with methods for the determination of the focal properties (excluding aberrations), aperture sizes, the field of view and resolving power of optical systems.

In all cases full descriptions are given, and the text is adequately supplied with diagrams. The references constitute an extensive bibliography of the subject. In short, this is an excellent book which will earn deserved praise from all those engaged in the production and testing of optical systems. The second volume will be awaited with interest. H. H. HOPKINS.

Mathematics of Engineering Systems (*Linear and Non-linear*), by D. F. LAWDEN. Pp. viii + 380. (London: Methuen; New York: John Wiley, 1954.) 30*s*.

'Engineering systems' is far too broad a category to be carried by the title of this book. In fact the book deals, from an electrical standpoint, with the mathematics of those electrical systems whose behaviour can be described by ordinary linear differential equations with constant coefficients under stated initial conditions. A final chapter provides an introduction to non-linear methods.

The author assumes that the reader has the calculus and complex numbers to first year engineering degree standard, and has an acquaintance with electric circuits and electrical engineering. Accordingly the mathematics is not presented abstractly but in terms of the physical quantities and the problem considered, so that it remains vivid and interesting. Clearly many years of teaching mathematics to engineers have contributed to the author's correct approach and his care in ploughing through, so as to clear the ground at points of real difficulty. Illustrated worked examples and the exercises set in each section of a chapter are also well chosen. The book comprises five chapters.

Chapter I is introductory, and as well as providing results required in later chapters is a useful revision course in the calculus and complex numbers to first year engineering degree level. The mathematics is reasonably rigorous, but it is a pity to note that the author refuses to grasp the nettle of the ϵ, δ definition of continuity or limit. Complex variable is treated in a simple manner as far as elementary functions and is just sufficient for the later applications. The student must look elsewhere for conformal mapping and contour integration.

Chapters II and III constitute the heart of the book. Here the author shows how to solve those electrical circuit problems giving rise to an equation of the form $\phi(D)x(t)=\psi(D)e(t)$, where ϕ and ψ are known polynomials in $D\equiv d/dt$, and $e(t)$ is a known input function. Chapter II presents the usual Complementary Function and Particular Integral method using D as an operator (Heaviside). The matter is illustrated, as elsewhere in the book, by quite a variety of worked examples of both practical and theoretical utility. Careful attention is given to the question of stability including the Hurwitz criteria. In Chapter III the treatment becomes more advanced, and by using the superposition principle (Duhamel) the responses of a linear system to unit input, unit impulse, and to a sinusoidal input are obtained. The effect of initial discontinuities is treated at length, by finding sufficient of the initial derivatives of the required output and then using these to establish the arbitrary constants in the complementary function. The method is most instructive, but laborious in practice when compared with the Laplace transform technique, as the author well recognizes. The Laplace transform method, however, is considered in this chapter as far as and including the convolution integral, without the use of the inversion theorem. Here a greater emphasis on the more powerful transform method and its relation to the classical approach would have proved more economical and supplied a better balance. The choice of the non-p-multiplied definition of the Laplace integral seems somewhat injudicious in a work where the operator D performs a large and essential role. Thus the close connection between the operational solution of $\phi(D)x(t)=H(t)$, viz. $x(t)=(1/\phi(D))\ H(t)$, and the p-multiplied transform solution $\bar{x}=1/\phi(p)$ is not made apparent. The Nyquist stability criterion is obtained in a careful and detailed treatment.

In Chapter IV the essential features of Fourier series, the Fourier integral and associated transforms are considered and applied to various problems such as the relation between harmonic response and that of unit function. Rigorous mathematics can no longer be expected, but the author is careful to state sufficient conditions for the validity of the results. Here again more cohesion and insight could have been achieved had the relation been made apparent between the Fourier and Laplace transforms.

The final Chapter V is a useful introduction to the more difficult subject of non-linear equations. The basic analytic techniques are presented with special reference to Van der Pol's equation and to Duffing's equation. A little more use could have been made of the phase plane, especially the role of the singular points, while no attempt is made to discuss stability.

The book can be recommended with confidence, particularly to those in the electrical and allied fields who require a textbook, intermediate in character between the elementary and advanced works, from which they can readily work on their own account. A. J. MAUTNER.

Abstracts of the Literature on Semiconducting and Luminescent Materials and their Applications (1953 issue), compiled by the BATTELLE MEMORIAL INSTITUTE. Pp. x+169. (New York: John Wiley; London: Chapman and Hall, 1955.) 40*s*.

The number of papers on semiconductors and related topics published each year is so large it is almost impossible for persons working in this field to be fully cognizant with all the important developments taking place. This book should be very valuable and the authors are to be congratulated on taking on such a noble task.

The volume under review covers the period January to December 1953. The progress in this field is so rapid that some of the references are already of historical interest only. It is to be hoped that the next volume will appear more rapidly now that a year's experience has been gained. There are 775 references and a short abstract is given of the majority. Classification is by Materials, covering the whole range from Germanium to Organic semiconductors. This seems by far the best classification system that could be adopted in the circumstances. In addition there is an author and subject index.

A very real attempt has been made to include every relevant publication and the coverage is excellent, including as it does unpublished papers read at various Society meetings in the U.S.A. and a few (American) patent specifications. The latter source is not, however, used to the extent it could be.

Some errors in abstracting are presumably inevitable in an undertaking of this sort and do not detract materially from the very real value of the work as a whole. Nevertheless, the accuracy of the abstracts is not up to the very high standard set by the established abstract journals. A. F. GIBSON.

Physics—A Descriptive Interpretation, by C. H. BACHMAN. Pp. viii + 497. (London: Chapman and Hall; New York: John Wiley, 1955.) 44*s*.

This book is intended to give students who are not specialists in science a connected account of what physics is about and what ideas physicists use in their investigations. The scheme of the book depends on logical convenience rather than on conventional classification—the initial chapters on basic concepts, energy, momentum and the electromagnetic spectrum lead rapidly to atomic and nuclear physics. This 'microcosmic' physics is followed by the 'macrocosmic' physics of heat, classical electricity and magnetism, geometric and wave optics. The second half of the book is concerned with the application of physical ideas to the world we know—there are long sections on the 'cosmic physics' of the weather, the earth, the stars and the galaxies and on the physics of biology, particularly of various processes in the human body. This is one of the most attractive sections of the book. There are final chapters on historic development, stressing the accelerating pace of scientific progress, and on the present structure of scientific research—the latter from a distinctly American point of view. The book is written without mathematics other than arithmetic and elementary algebraic formulae. There are a number of sets of discussion topics; many of them would probably prove unreasonably difficult to a person with no other guide than the text. The aims and arrangement of the book are original and commendable: it was obviously written with American college courses in mind, but in this country might find its greatest appreciation in the sixth forms of schools and in adult education classes, for both of which it is eminently suitable.

The reviewer detects no serious inaccuracies or misleading treatments though a few of the analogies seemed to him pointless and laboured. A few of the drawings seemed ill-chosen and some of them showed a comic-strip facetiousness which might repel the staider reader. H. J. J. BRADDICK.

Light Calculations and Measurements, by H. A. E. KEITZ. Pp. xvi + 413. (Eindhoven: Philips Technical Library; London: Cleaver Hume Press, 1955.) 52*s*. 6*d*.

This book is essentially a textbook on photometry in which the basic concepts are considered in Part I under the somewhat misleading title of Light Calculations, while Part II covers the more practical side of the subject under the general heading of The Measurement of Light. It is a good book, excellently produced, the subject is at most points approached in an orthodox manner, and the treatment is clear and generally very sound indeed.

Part I is on the whole more exhaustive than Part II, for the discussion of physical photometers is rather slight. Then since the photometry of optical systems is considered in Part I, it is a pity that more account is not given in Part II of the measurement of the light transmission or the veiling glare of an optical system. No reference, for example, is given to the methods described by Guild and used at the National Physical Laboratory, or to the equivalent methods in use elsewhere. Another omission, both as regards principle and practice, is the photometry of flashing light sources.

Heterochromatic photometry is quite correctly described as based on the international relative luminous efficiency function V_λ and it is perhaps inadvisable in a textbook of this kind to hint that in some quarters at least this cornerstone of photometry is believed to be working loose. Summation of photometric quantities is also accepted, although it is recognized that equality of luminance does not necessarily imply equivalence in all respects. Whether the statement that far higher visual acuity is found in sodium light than in tungsten light of the same photometric value is either correct or a good example of the point at issue is, however, open to question.

But these are, in the main, very minor criticisms and the author is to be congratulated on a fine book. Although it was written primarily for lighting engineers, it may confidently be recommended as a textbook on photometry suitable for undergraduate physics students and as a practical reference volume suitable for physicists who encounter problems of light measurement in their research. W. D. WRIGHT.

BOOK NOTICES

Relativity for the Layman (a simplified account of the history, theory and proofs of relativity), by JAMES A. COLEMAN. Pp. 131. (New York: William-Frederick Press, 1954.) $ 2.75.

Handbook of the [Science Museum] *Collection illustrating Temperature Measurement and Control. Part II. Catalogue of Exhibits, with Descriptive Notes,* by J. A. CHALDECOTT. Ministry of Education, the Science Museum. (London: H.M. Stationery Office, 1955.) 3*s.* 6*d.*

Massbalancing of Aircraft Control Surfaces, by H. TEMPLETON. Royal Aeronautical Society, Aeronautical Monographs No. 4. Pp. x + 241. (London: Chapman and Hall, 1954.) 35*s.*

Étude des textures cristallines obtenues par laminage progressif de monocristaux d'aluminium, by JACK MANENC. Pp. 117. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, 1954.) 1400 *fr.*

Étude de frottement interne par élimination systematique des forces d'inertie, by A. BRODEAU. Pp. 15. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, N.T. 54, 1955.) 300 *fr.*

Structure cristalline de la cancrinite, by P. NITHOLLON. Pp. 48. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, N.T. 53, 1955.) 425 *fr.*

Comparaison de la résistance au choc des films lubrifiants, by F. CHARRON. Pp. 18. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, No. 298, 1955.) 400 *fr.*

Méthodes d'algèbre abstraite en géométrie algébrique, by P. SAMUEL. Pp. x + 133. (Berlin, Göttingen, Heidelberg: Springer 1955.) (Ergebnisse der Mathematik und ihrer Grenzgebiete, Neue Folge, Heft 4.) DM 23.60.

L'exploitation des données empiriques. Actes du colloque tenu à Luxembourg en juillet 1953 dans le cadre du Congrès de l'Association Française pour l'Avancement des Sciences. Pp. 77. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, N.T. 52, 1955.) 850 *fr.*

L'interpolation idéale et son calcul numérique effectif, by P. Vernotte. Pp. iii + 74. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, No. 300.) 850 *fr.*

Sensitométrie photographique, by M. Roulleau. Pp. 194. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, No. 301, 1955.) 1400 *fr.*

Nouveaux compléments d'hydraulique, deuxième partie, by L. Escande. Pp. 274. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, No. 302, 1955.) 2400 *fr.*

Mouvement oscillatoire avec viscosité et inertie, by C. Clarion. Pp. 76. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, No. 303, 1955.) 1000 *fr.*

Etude thermique des écoulements d'air raréfié en régime moléculaire libre, by F. M. Devienne. Pp. 31. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, N.T. 55, 1955.) 400 *fr.*

Zur Ermittlung von Unterschallströmungen mit der Transformationsmethode bei quadratischer Approximation der Adiabate, by H. Schubert and E. Schineke. Pp. 32. (Berlin: Akademie-Verlag, Berichte über die Verhandlungen der Sächsischen Akademie der Wissenschaften zu Leipzig, Mathematisch-naturwissenschaftliche Klasse, Band 101, Heft 6, 1955.) DM 2.30.

CONTENTS OF SECTION A

PAGE

Editorial 1

Mr. J. H. **Haywood**. The Equations of Motion of Rotating Bodies in General Relativity 2

Dr. T. S. **Green** and Dr. R. **Middleton**. An Investigation of Deuteron Induced Reactions by Magnetic Analysis—I : Experimental Technique 16

Dr. T. S. **Green** and Dr. R. **Middleton**. Investigation of Deuteron Induced Reactions by Magnetic Analysis—II : Results for ^{9}Be, ^{12}C, ^{14}N and ^{16}O . 28

Mrs. G. A. **Chackett**, Dr. K. F. **Chackett**, Dr. P. **Reasbeck**, Dr. J. L. **Symonds** and Mr. J. **Warren**. Excitation Functions up to 980 mev for Proton-Induced Reactions in ^{27}Al, relative to ^{12}C(p, pn)^{11}C 43

Dr. A. C. **Hurley**. On the Method of Atoms in Molecules—II : An Intra-Atomic Correlation Correction 49

Dr. A. **Dalgarno** and Dr. J. T. **Lewis**. The Representation of Long Range Forces by Series Expansions—I : The Divergence of the Series; II : The Complete Perturbation Calculation of Long Range Forces 57

Dr. B. H. **Bransden** and Dr. A. **Dalgarno**. The Calculation of Auto-Ionization Probabilities—III : Auto-detachment from the Negative Hydrogen Ion . 65

Research Notes :

Mr. R. K. **Gupta** and Dr. S. **Jha**. A High Energy Gamma Ray in the Decay of Barium 131 70

Reviews of Books 72

Contents of Section B 76

Principal Lattice Frequency of MgO

By B. D. SAKSENA and S. VISWANATHAN
National Physical Laboratory of India, New Delhi, India

MS. received 26th *January* 1955 *and in amended form* 13th *September* 1955

Abstract. The principal lattice frequency has been obtained at 25·26 μ from the observed dispersion in the region 0·36–5 μ. This frequency and the frequencies of the longitudinal and transverse modes of vibration of (1, 0, 0) and (1, 1, 1) planes against each other have been calculated on the Born model.

Reflection coefficients and refractive indices have been calculated, in good agreement with observed data, on the usual dispersion formula and that of Blackman. The lattice frequency lies at 25·1 μ in absorption and at 22 μ in reflection, the reflection curve between 17 and 22 μ being flat. The reflection maximum at 16 μ is interpreted as a combination frequency.

§ 1. Introduction

Magnesium oxide is a cubic crystal with NaCl structure and there is therefore only one principal lattice oscillation which is active only in the infra-red.

The infra-red spectrum of MgO has been investigated by a number of workers but this frequency is still undecided. Tolksdorf (1928) observed a strong absorption at 14·2 μ; Strong (1931), using reststrahlen, found a strong absorption in the neighbourhood of 23 μ. Fock (1934) found a strong absorption maximum at 17·3 μ with minor maxima at 14·9 μ and 25 μ. Barnes, Brattain and Seitz (1935) were unable to identify the fundamental absorption band though they obtained two sharp reflection peaks 14·82 and 15·3 μ. Burstein, Oberley and Plyler (1948) find a very broad band in reflection with a major peak near 22 μ and a minor one near 16 μ. Willmott (1950) also finds the reflection maximum to be very broad extending from 17 to 21 μ. He considers that the eigenfrequency is at 17·3 μ and subsidiary absorption maxima at 14·8 μ and 24·3 μ. Momin (1953) has investigated the region between 1 and 21 μ and finds the strongest absorption band between 18 and 19 μ.

In the luminescence spectra of MgO(Cr) phosphors, Deutschbein (1932) observed a number of bands on either side of a sharp line at 6991 Å which he interpreted as being due to the combination of the electronic transition at 6991 Å (of Cr^{3+} ions) with the vibrational frequency of the MgO lattice. In this way he identified a frequency at 425 cm^{-1} as the lattice frequency.

Accurate data regarding the refractive index of MgO in the region 0·36 to 5·3 μ are available from the work of Stephens and Malikson (1952). We have used this data to obtain the value of the lattice frequency and have also obtained its position in reflection and absorption from the infra-red studies of Willmott and of Burstein, Oberley and Plyler.

§ 2. CALCULATION OF THE LATTICE FREQUENCY FROM THE DISPERSION DATA

Stephens and Malikson give the following expression for n^2 where n is the refractive index:

$$n^2 = 2{\cdot}956362 - (0{\cdot}01062387\lambda^2 + 0{\cdot}0000204968\lambda^4) + \frac{0{\cdot}0219577}{\lambda^2 - 0{\cdot}01428322}. \quad \ldots\ldots(1)$$

The expression in brackets involving λ^2 and λ^4 can be put in the form $a\,\lambda^2/(\lambda_l^2 - \lambda^2)$ with $a = 5{\cdot}52$ and $\lambda_l = 22{\cdot}7\,\mu$. This gives a value 8·47 for the static dielectric constant (n_∞^2), the value obtained by Højendahl (1938) being 9·8. Using the latter value we get $a = 6{\cdot}85$. The value of λ_l may be calculated from expression (6) of the next section with $r = 0$. Using $\lambda_l = 25{\cdot}26\,\mu$ and $a = 6{\cdot}85$, the observed and calculated values of n in the entire region differ by about 0·0001 only in a few cases (see table 1).

Table 1

λ (μ)	0·36117	0·36501	0·40465	0·43583	0·48613	0·54607	0·58926
n_{obs}	1·77318	1·77186	1·76099	1·75465	1·74706	1·74073	1·73722
n_{cal}	1·77313	1·77186	1·76111	1·75480	1·74720	1·74083	1·73740
λ (μ)	0·65628	0·66781	0·70652	0·76786	1·01398	1·12866	1·36728
n_{obs}	1·73330	1·73272	1·73098	1·72868	1·72259	1·72059	1·71715
n_{cal}	1·73333	1·73275	1·73089	1·72864	1·72249	1·72049	1·71699
λ (μ)	1·52952	1·6932	1·7092	1·81307	1·97009	2·24929	2·32542
n_{obs}	1·71496	1·71281	1·71258	1·71108	1·70885	1·70470	1·70350
n_{cal}	1·71482	1·71264	1·71243	1·71102	1·70883	1·70469	1·70349
λ (μ)	3·3033	3·5078	4·258	5·138	5·35		
n_{obs}	1·68526	1·68055	1·66039	1·63138	1·62404		
n_{cal}	1·68499	1·68031	1·66041	1·63143	1·62323		

We are interested in the dispersion formula for the infra-red region. At $\lambda = \lambda_l$, the last term in expression (1) gives a small contribution of 0·000033 to n^2, and therefore in the neighbourhood of λ_l the dispersion may be written as

$$n^2 = 2{\cdot}956 + a\lambda^2/(\lambda^2 - \lambda_l^2) \quad \ldots\ldots(2)$$

$$= q + C_l/(\omega_l^2 - \omega^2) \quad \ldots\ldots(3)$$

$$\text{where } \omega = 2\pi c/\lambda,\ q = 2{\cdot}956 \text{ and } a = C_l\lambda_l^2/4\pi^2c^2. \quad \ldots\ldots(4)$$

Relation (2) is a Drude type dispersion formula and so we can regard ω_l as the observed resonance frequency (Krishnan and Roy 1953). Considering that this is the principal lattice oscillation, its oscillator strength will be defined by Born's relation (1933)

$$C_l = 4\pi Ne^2/m \quad \ldots\ldots(5)$$

where m is the reduced mass of the Mg and O ions, N the number of ion pairs per unit volume, e the charge on the ions which is here equal to twice the electronic charge. The values of C_l calculated from (4) and (5) agree closely, being $381{\cdot}0 \times 10^{26}$ and $389{\cdot}7 \times 10^{26}$ respectively.

Expression (3) can be used to find the refractive index below $13{\cdot}87\,\mu$ and above $30\,\mu$. The values of n calculated by this formula from 6 to $13\,\mu$ are shown in figure 1 and agree with the data of Willmott. We find that n diminishes from a value close to unity at $12\,\mu$ to a very low value at $14\,\mu$ which confirms the values of constants a and λ_l used here. Also, in agreement with Strong we find that $n = 4{\cdot}23$ at $35\,\mu$. At $13{\cdot}87\,\mu$, n^2 becomes zero and beyond it negative. Therefore, between 13·87 and $30\,\mu$ we have to use a modified expression such as (10) or (15).

§ 3. CALCULATION OF THE LATTICE FREQUENCY ON BORN'S THEORY

According to Mott and Littleton (1938), the resonance frequency ν_l of a cubic crystal is given by

$$\frac{\epsilon - n_0^2}{[1+\frac{1}{3}r(n_0^2-1)]^2} = \frac{Ne^2}{m\pi\nu_l^2} \qquad \ldots\ldots(6)$$

where ϵ is the static dielectric constant and n_0 the refractive index extrapolated to zero wavelength. Since the dispersion of MgO is given by the Drude formula, the overlap factor r (for $r=1$ there is no overlap between the ions) must be zero.

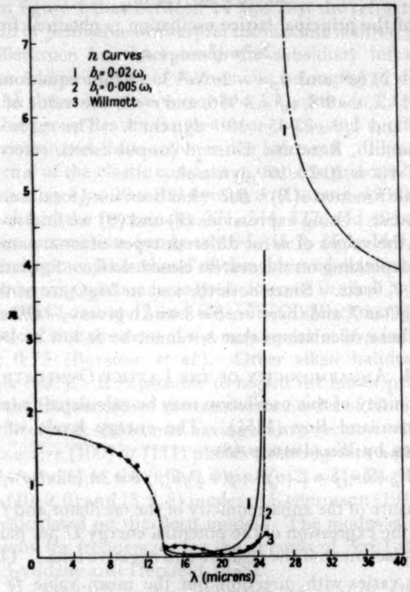


Figure 1.

b_1 in this diagram should read b.

The polarization of the medium is therefore, wholly due to the relative displacement of positive and negative ions. Krishnan and Roy (1951) have found similar results in alkali halides. Now although MgO differs from alkali halides in that c_{12} is not equal to c_{44}, it may be instructive, as a first approximation, to calculate the principal lattice frequency by their method. They assume, following Born, that the ions are held in their respective positions by electrostatic forces between the charges and by short range forces of repulsion between the nearest neighbours. Taking the energy of repulsion between neighbouring ions separated by a distance R as $\phi(R) = Ae^{-R\rho}$ where A and ρ are constants, the potential energy U of the crystal per pair of ions is given by

$$U = -\alpha e^2/R + p_0 A e^{-R\rho}$$

where α is the Madelung constant and p_0 the number of nearest neighbours.

The values of A and ρ can be evaluated from the relations

$$\left(\frac{dU}{dR}\right)_{R=d}=0 \quad \text{and} \quad \left(\frac{d^2U}{dR^2}\right)_{R=d}=\frac{9}{N\beta d^2} \qquad \ldots\ldots(7)$$

where β is the isothermal compressibility of the crystal, N the number of ion pairs per unit volume, and d the value of R when the crystal is in equilibrium. Both d and β appearing in these expressions refer to absolute zero of temperature. Assuming $d/\rho=\delta$ we obtain from (7)

$$\delta-2=9d/N\alpha e^2\beta. \qquad \ldots\ldots(8)$$

The frequency of the principal lattice oscillation is obtained from the relation

$$2\pi^2 m\nu_1{}^2=a_1+a_2 \qquad \ldots\ldots(9)$$

where $a_1=\alpha e^2(\delta-2)/6d^3$ and $a_2=-2\pi Ne^2/3$. From equations (9) and (6) we get, using $d=2{\cdot}11$ Å, $\epsilon=9{\cdot}8$, $n_0{}^2=2{\cdot}956$, and $r=0$, the value of λ_l to be equal to $25{\cdot}26\,\mu$, $\delta=7{\cdot}17$ and $1/\beta=23{\cdot}35\times10^{11}$ dyn cm^{-2}. The value of $1/\beta$ according to Quimby, Balamuth, Rose and Durand (unpublished, referred to by Barnes *et al.* 1935) at 77·4°K is $20{\cdot}2\times10^{11}$ dyn/cm^2.

Another simple function $\phi(R)=BR^{-m}$ has been used for the energy of repulsion between neighbours. Using expressions (7) and (8) we find $m=\delta-1$. Pauling (1928) has given the values of m for different types of ions, namely helium, neon and argon type (depending on the nearest closed shell configuration to which they correspond) as 5, 7, 9, etc. Since both the ions in MgO are of the neon type, the value of m for MgO is 7, and therefore $\delta=8$ which gives $\nu_l=493$ cm^{-1} ($\lambda_l=20{\cdot}3\,\mu$). It follows from these calculations that λ_l cannot lie as low as $14{\cdot}2\,\mu$ or $17\,\mu$.

§ 4. ANHARMONICITY OF THE LATTICE OSCILLATION

The anharmonicity of this oscillation may be calculated from the expressions given by Krishnan and Roy (1951). The energy levels of an anharmonic oscillator are given by Wu (1946, p.48)

$$W_p=h\nu_1[p+\tfrac{1}{2}+(p^2+p+\tfrac{1}{2})f\theta], \quad \theta=3h/(32\pi^4m^2\nu_1{}^3)$$

where $f\theta$ is a measure of the anharmonicity of the oscillator and f is the coefficient of the r^4 term in the expression of the potential energy U per pair of ions where r is the relative displacement of the positive and negative ions. The anharmonicity of the oscillation varies with direction but the mean value $\overline{f\theta}$ is found to be $0{\cdot}8\times10^{-2}$ and the average value of f (i.e. $\overline{f}$) is $5{\cdot}29\times10^{20}$ which is of the same order as the anharmonicity of the principal lattice oscillation in alkali halides. The anharmonicity of the vibrations has pronounced influence on the specific heat at constant volume at high temperatures (Born and Brody 1921); but we need to know the anharmonicities of all the other normal modes of the crystal for this purpose. We have no data on the variation of C_v with temperature at high temperatures for MgO. Present calculations suggest that C_v should decrease by 1% for nearly 180° rise of temperature.

§ 5. POSITION OF SUBSIDIARY MAXIMA

We have found that the principal lattice oscillation occurs at $25{\cdot}26\,\mu$, and therefore all the other maxima obtained by Burstein, Oberley and Plyler and by Willmott should be treated as subsidiary maxima. Both Willmott and Burstein *et al.* find two secondary absorption peaks at 10·5 and $11{\cdot}8\,\mu$ in transmission. Besides these we have to treat the maxima at 14·8 and $17{\cdot}3\,\mu$ observed by Willmott as subsidiary maxima. It is probable that the broad maximum at $17\cdot3\,\mu$ is the same as the broad band at $16\,\mu$ observed by Burstein and others in reflection. The following arguments support the suggestions made in this paragraph: (i) All the maxima lie on the short wavelength side of the eigenfrequency at $25\cdot26\,\mu$. This is consistent with the absorption spectra of other face-centred cubic crystals where a similar result has been obtained (Parodi 1937, 1938). (ii) According to Blackman (1937), the secondary maxima on the short wavelength side of the main maximum fall usually in a region where the refractive index is very small. We may see from figure 1 that between 14 and $20\,\mu$ the refractive index is very small, and it would be justifiable to interpret the maxima in this region as secondary maxima. (iii) Blackman has interpreted the subsidiary infra-red maxima in NaCl and KCl as additive tones of frequencies belonging to different branches but associated with the same wave vector $k(=1/\lambda)$ of the $(\pi/2, \pi/2, \pi/2)$ group of frequencies. This group has six frequencies ν_1, ν_2, ν_3 and ν_4 (the first two being double) whose combinations are active in the infra-red. These frequencies are expressible in terms of the elastic constant c_{11} and the principal lattice frequency ν_1. For MgO (using $c_{11}=29\times10^{11}$ dyn cm^{-2} (Durand 1936), $d=2\cdot11$ Å, and $\lambda_1=25\cdot26\,\mu$) we obtain the frequencies 271, 344, 327, and 404 cm^{-1} whose combinations give six summation bands at 16·5, 15·7, 15·1, 16·7, 13·6 and $14\cdot8\,\mu$ which are in the close neighbourhood of the observed maxima at 14·8 and $16\,\mu$. Similar closely occurring maxima have been observed in alkali halides, for example at 40 and $50\,\mu$ in NaCl and at 47 and $60\,\mu$ in KCl (Blackman 1937).

It may be noticed that the ratio of the minor peak at $16\,\mu$ in MgO to the major peak at $22\,\mu$ is 0·73 (Burstein *et al.*). Other alkali halides show a similar behaviour (Hohls 1937). It is possible to regard the minor peak as due to summation tones of the frequencies of the transverse modes of vibration of (100) or of (111) planes of the crystal, the former having a wave vector $k=1/2d$, and the latter $\sqrt{3}/4d$, (the successive (100) or (111) planes vibrating with a phase difference π). These modes correspond to the $(\pi/2, 0, 0)$ and $(\pi/2, \pi/2, \pi/2)$ groups of Blackman or to the $p_{x/y/z}=(10, 0, 0)$ and $(5, 5, 5)$ modes of Kellermann (1940). The frequencies have been calculated on the Born model. The modes of vibration and the explicit expressions for frequencies for the crystals of NaCl structure are given below (see also Lyddane and Herzfeld 1938).

(1, 0, 0) planes, longitudinal modes: vibration of (1, 0, 0) planes against each other along the x axis

$$[\{\alpha(\delta-2)/6+1\cdot08\}-2\pi^2(d^3/e^2)m_{\text{cat}}\nu^2][\{\alpha(\delta-2)/6+1\cdot08\}-2\pi^2(d^3/e^2)m_{\text{an}}\nu^2]$$
$$=\{\alpha(\delta+2)/6-3\cdot74\}^2$$

(1, 0, 0) planes, transverse modes: (1, 0, 0) planes vibrating against each other along y or z axis

$$[\{\alpha(\delta-2)/6-0\cdot54\}-2\pi^2(d^3/e^2)m_{\text{cat}}\nu^2][\{\alpha(\delta-2)/6-0\cdot54\}-2\pi^2(d^3/e^2)m_{\text{an}}\nu^2]$$
$$=\{\alpha\delta/6-1\cdot87\}^2$$

(1, 1, 1) planes, longitudinal modes: vibration of (1, 1, 1) planes of anions only or of cations only, along cube diagonal

$$(e^2/d^3)[\alpha(\delta-2)/6+1\cdot808]=2\pi^2\nu^2 m_{\text{an/cat}}$$

(1, 1, 1) planes, transverse modes: vibration of (1, 1, 1) planes of anions alone or of cations alone perpendicular to the cube diagonal

$$(e^2/d^3)[\alpha(\delta-2)/6-0\cdot904]=2\pi^2\nu^2 m_{\text{an/cat}}.$$

For MgO we use $\delta = 7{\cdot}17$ (see § 3), and for the alkali halides the value given by Krishnan and Roy (1951). The frequencies, expressed in cm^{-1}, have been tabulated in table 2.

Table 2

	(1)	(2)				(3)						(4)
		(*a*)		(*b*)		(*a*)			(*b*)			
		opt.	ac.	$\bar{\nu}_{\mathrm{an}}$	$\bar{\nu}_{\mathrm{cat}}$	opt.	ac.	sum	$\bar{\nu}_{\mathrm{an}}$	$\bar{\nu}_{\mathrm{cat}}$	sum	
LiF	385	618	373	670	409	358	209	567	276	167	443	588
NaF	278	334	243	315	347	225	149	374	152	167	319	400
NaCl	163	224	170	243	196	171	94	265	132	107	239	256
		(223	164	240	193	164	94	—	127	102)	—	—
KCl	140	158	146	164	172	139	71	210	96	100	196	218
NaBr	140	223	119	227	122	153	62	215	129	69	198	—
KBr	113	142	99	156	109	116	52	168	94	66	160	—
MgO	396	800	526	673	830	462	340	802	286	354	640	625

(1) Reststrahlen frequency $\bar{\nu}_1$ (Barnes 1932); (2) $\bar{\nu}$ longitudinal modes, (*a*) (100) plane, optical, acoustic, (*b*) (111) plane, $\bar{\nu}_{\mathrm{an}}$, $\bar{\nu}_{\mathrm{cat}}$; (3) $\bar{\nu}$ transverse modes, (*a*) (100) plane, optical, acoustic, sum, (*b*) (111) plane, $\bar{\nu}_{\mathrm{an}}$, $\bar{\nu}_{\mathrm{cat}}$, sum; (4) observed infra-red maxima.

It may be seen that the sum of the frequencies of the two transverse modes in these vibrations is roughly $\sqrt{2}\nu_1$. This relation appears to be fairly accurate for NaCl, KCl, NaBr, and KBr. In the Raman spectra of these crystals, Menzies and Skinner (1948) observed depolarized bands at 232, 122, 154, and 128 cm^{-1} respectively. It is quite likely that these frequencies correspond to the same mode of vibration. Since in the Raman effect the octave 2ν, where ν is the frequency of the normal mode, is active, the frequency of the given mode in the above crystals is 116, 107, 77, 64 cm^{-1} respectively. These values are very close to the values of $(\bar{\nu}_{\mathrm{cat}})_{\mathrm{tr}}$ (107, 100, 69, 66) in the spectra of these crystals. It would be interesting to find if the Raman spectra of MgO shows a depolarized band at 708 cm^{-1}.

It is possible that the subsidiary maxima at $10{\cdot}5$ and $11{\cdot}8\,\mu$ in MgO may be due to the combination of one transverse and one longitudinal mode belonging to the same wave vector. In this way we get summation bands at $10{\cdot}1$, $10{\cdot}4$, and $11{\cdot}6\,\mu$. The corresponding summational modes for NaCl can be calculated as $37{\cdot}9$, $29{\cdot}4$ and $27{\cdot}0\,\mu$. It is interesting to note that maxima close to $30\,\mu$ and $40\,\mu$ have been observed by Mentzel (1934) for NaCl.

The frequencies given in parenthesis in table 2 for NaCl are those given by Kellermann for $p_{x/y/z} = (10, 0, 0)$ and $(5, 5, 5)$ modes.

§ 6. Optical Properties of MgO in the Infra-red

The dispersion in the neighbourhood of a resonance frequency may be written as (Havelock 1912)

$$n'^2 = n_0{}^2 + C_1/[(\omega_1{}^2 - \omega^2) + jb\omega] \qquad \ldots\ldots(10)$$

where $n' = n(1 - jk)$ is the complex refractive index. ......(11)

The term b in expression (10) is called the damping term. We have

$$n^2(1 - k^2) = n_0{}^2 + C_1(\omega_1{}^2 - \omega^2)/[(\omega_1{}^2 - \omega^2)^2 + b^2\omega^2] \qquad \ldots\ldots(12a)$$

$$2n^2k = C_1 b\omega/[(\omega_1{}^2 - \omega^2)^2 + b^2\omega^2]. \qquad \ldots\ldots(12b)$$

Here ω is the angular frequency $= 2\pi c/\lambda$. From equation (12) we can get n and nk for any wavelength λ if b is known. The reflection coefficient

$$R = [(n-1)^2 + n^2k^2]/[(n+1)^2 + n^2k^2].$$

When ω is very far removed from ω_1, the term $jb\omega$ can be neglected and expression (10) reduces to (3).

We may ask whether b is a constant or varies with the frequency ω of the electromagnetic wave traversing the crystal. Since damping results from the frictional resistance offered by the medium to the vibrations, and also from the electromagnetic radiation damping, the damping factor is by definition a constant for a given oscillator. However, if there is a large number of simple harmonic oscillators in a medium, the damping factor may be altered due to mutual interactions. We do not know theoretically how b should vary with ω, and so it is worth while to examine the experimental data in this connection.

Since nk is the imaginary part of the refractive index in expression (11), the absorption coefficient μ_s is given by

$$\mu_s = -(1/s)\log(I/I_0) = 4\pi nk/\lambda \qquad \ldots\ldots(13)$$

where I is the intensity of the beam of light of wavelength λ after it has traversed a length s of the crystal. When we are far away from the resonance frequency, say, between 6 and 12 μ, the relation (12*b*) can be written without any appreciable error as

$$2n^2k = C_1 b\omega/[(\omega_1{}^2 - \omega^2)^2] \qquad \ldots\ldots(14)$$

while relation (12*a*) reduces to expression (3) and gives n directly for various wavelengths. Knowing μ_s one can study the variation of b with wavelength with the help of relations (13) and (14). Table 3 gives the results for the region 6 to 13μ.

Table 3

		$nk \times 10^3$				b/ω_1	
λ (μ)	n	B.O.P.	B.B.S.	W	Mean	from (13) & (14)	from (15)
6·0	1·595	0·05	—	—	0·05	0·0015	0·007
7·0	1·545	0·117	—	—	0·117	0·0020	0·007
8·0	1·487	0·35	—	—	0·35	0·0038	0·012
9·0	1·400	0·96	0·58	0·58	0·71	0·0050	0·014
9·25	1·377	—	—	0·87	0·87	0·0053	0·014
9·50	1·358	—	0·91	1·33	1·12	0·0061	0·016
9·75	1·325	—	—	2·22	2·22	0·0108	0·028
10·00	1·297	5·25	3·29	4·26	4·27	0·0180	0·045
10·25	1·268	—	—	3·88	3·88	0·0150	0·037
10·50	1·235	—	4·29	3·52	3·95	0·0140	0·033
10·75	1·200	—	—	5·20	5·20	0·0160	0·037
11·00	1·163	14·6	4·54	8·43	9·19	0·0250	0·057
11·25	1·124	—	—	15·70	15·70	0·0370	0·083
12·00	0·986	—	—	15·90	15·90	0·0250	0·052
12·50	0·857	—	13·9	—	13·90	0·0160	0·032
13·00	0·699	35·0	—	—	35·00	0·0290	0·056

B.O.P., Burstein, Oberley and Plyler; B.B.S., Barnes, Brattain and Seitz; W, Willmott.

Except for the wavelengths 10·0 and 11·25 μ where there is additional absorption due to subsidiary maxima, it is obvious that the value of b (see 7th column), instead of being a constant, increases nearly 20 times (from 0·0015 ω_1 to 0·029 ω_1) between 6 and 13 μ.

Blackman considers that, for a polar crystal, the expression $jb\omega$ in expression (10) should be replaced by $jb\omega_1$. We then have

$$n'^2 = n_0^2 + C_1/[(\omega_1^2 - \omega^2) + jb\omega_1] \qquad \ldots\ldots(15a)$$

$$2n^2k = C_1 b\omega_1/[(\omega_1^2 - \omega^2)^2 + b^2\omega_1^2] \qquad \ldots\ldots(15b)$$

The difference between (12) and (15) lies in the fact that, whereas for (12) nk tends to zero as ω tends to zero, according to (15) nk tends to a constant value $(C_1/2\sqrt{\epsilon})b/\omega_1^3$ as ω tends to zero. This means that the absorption coefficient cannot tend to zero as long as b is finite; in other words we can have combination vibrations which can absorb energy. We have also calculated b/ω_1 with the help of equation (15) using the experimental values of nk, and these values have been entered in the last column. We find that there is still a variation of b/ω_1 between 6 and 13 μ though the variation is smaller (being only 8 times and not 20 times as found previously).

We have found the values of the reflection coefficient R and refractive index n in the entire resonance region from 12 to 35 μ with b constant, (i) with the help of dispersion formula (10) and (ii) with the help of dispersion formula (15). For comparison we have the values of the reflection coefficients due to Burstein and others between 12 and 35 μ and also due to Willmott between 14 and 25 μ. The latter has also given the values of the refractive index between 7 and 24·5 μ. There are, however, significant differences in the data of these workers regarding the value of reflection maximum and the reflection coefficients in the region 25 to 35 μ.

Case (i): We have used two values for b, viz. $b = 0{\cdot}02\omega_1$ and $b = 0{\cdot}005\omega_1$ and have calculated the values of n (figure 1) and R and nk (figure 2). There is good agreement between the calculated value of R (for both the values of b) with those observed by Burstein and others between 25 to 35 μ. The reflection maximum is flat and lies between 18 and 22 μ. The value of $b = 0{\cdot}005\omega_1$ gives lower values of n and higher reflection coefficients (99%) in the region of reflection maximum than the other value. Also it agrees better with the refractive index data between 20 and 24·2 μ though there are deviations between 14 and 18 μ. On the other hand, the value $b = 0{\cdot}02\omega_1$ agrees better with the values of refractive index and reflection coefficients in the region 14 to 18 μ but it gives a reflection maximum of 96% and higher values of n between 18 and 25 μ.

It appears from these curves that the principal lattice frequency occurs close to 25·1 μ in absorption and close to 22 μ in reflection, the curve between 17 and 22 μ being very flat. The flatness of the reflection curve is due to the fact that in this region the refractive index n has a very low value. Willmott reports an absorption maximum which he calls a subsidiary maximum at 24·3 μ but Fock reports the position of this maximum to be 25 μ which is very close to the value found here.

The maximum of nk occurs slightly towards the short wavelength side of the resonance frequency while the maximum of n occurs slightly towards the longer wavelength side. The minimum of n occurs at 16 μ but according to Willmott its position is 20 μ.

It may be seen from figure 1 that whereas the calculated refractive index curves run almost straight (with a weak minimum at 16 μ) between 14 and 20 μ, the curve given by Willmott shows a small maximum with two minima one on either side. This is probably due to the presence of subsidiary maxima in this region.

Case (ii): We have used $b = 0{\cdot}05\ \omega_1$. The reflection coefficients between 12 and 18 μ and beyond 26 μ agree with the data of Willmott and of Burstein and others, the values of R at 14, 16 and 18 μ being 0·66, 0·91 and 0·92 in good agreement with the values observed. The reflection maximum has a value of 0·93 and occurs close to 20 μ the curve being almost flat between 18 and 22 μ. The refractive index has a value of 0·11 at 14 μ, drops to a minimum value 0·028 at 18 μ and then increases to 0·16 at 20 μ and 0·38 at 22 μ.

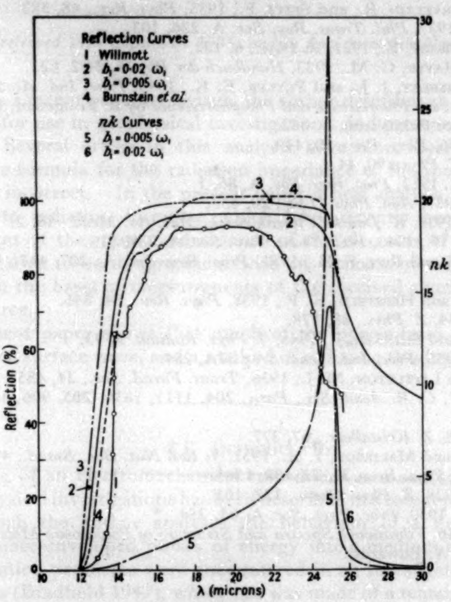


Figure 2.

b_1 in this diagram should read b.

For the region 6 to 13 μ it is possible to write empirically an expression for b which gives the same variation with ω as shown in the 7th column of table 3, but due to the limited range it cannot be considered to hold for the entire resonance region from 12 to 35 μ. The expression we have used is

$$b = A\omega_1/[(\omega/\omega_1 - \omega_1/\omega)^{2p} + B(\omega/\omega_1 + \omega_1/\omega)^q]$$

with $A = B = 2/3$ and $2p = q = 4$. In general the results are similar to case (i), but the values of refractive indices are much higher and the reflection maximum is 90%. The minimum of n occurs at the same place, i.e. 16 μ.

It appears from this study that the entire reflection curve from 12 to 18 μ and beyond 26 μ due to Willmott and Burstein *et al.* is consistent with a value of b greater than $0{\cdot}02\ \omega_1$, but the values of n obtained by Willmott for the region 18 to 22 μ are very much lower than those expected. The reflection maximum is flat and the value consistent with the above value of b is 93% obtained by Willmott and not 103% as found by Burstein and others. As already stated the absorption maximum is close to $25\,\mu$ and reflection maximum close to $21\,\mu$. The dispersion formula (15), due to Blackman, appears to fit the experimental results better than the dispersion formula (10) because it gives a minimum for n which is nearer to the experimental value of Willmott; also the variation of b with ω is much smaller.

REFERENCES

BARNES, R. B., 1932, *Z. Phys.*, **75,** 723.

BARNES, R. B., BRATTAIN, R., and SEITZ, F., 1935, *Phys. Rev.*, **48,** 582.

BLACKMAN, M., 1937, *Phil. Trans. Roy. Soc.* A, **236,** 103.

BORN, M., and BRODY, E., 1921, *Z. Phys.*, **6,** 132.

BORN, M., and MAYER, G. M., 1933, *Handbuch der Physik*, **24**/2, 623.

BURSTEIN, E., OBERLEY, J. J., and PLYLER, E. K., 1948, *Proc. Ind. Acad. Sci.*, **28A,** 398.

DEUTSCHBEIN, O., see *Fluorescence and Phosphorescence* by P. Pringstein, 1949, (New York : Interscience Publishers), p. 642.

DURAND, M., 1936, *Phys. Rev.*, **50,** 449.

FOCK, J., 1934, *Z. Phys.*, **90,** 44.

HAVELOCK, T. H., 1912, *Proc. Roy. Soc.* A, **86,** 1.

HOHLS, H. W., 1937, *Ann. Phys., Lpz.*, **29,** 433.

HØJENDHAL, K., 1938, *K. Danske Vidensk Selsk, Mat. fys. Medd.*, **16,** 2.

KELLERMANN, E. W., 1940, *Phil. Trans. Roy. Soc.* A, **238,** 513.

KRISHNAN, K. S., and ROY, S. K., 1951, *Proc. Roy. Soc.* A, **207,** 447; 1953, *Phil. Mag.*, **44,** 90.

LYDDANE, R. H., and HERZFELD, K. F., 1938, *Phys. Rev.*, **54,** 846.

MENTZEL, A., 1934, *Z. Phys.*, **88,** 178.

MENZIES, A. C., and SKINNER, J., 1948, *J. Phys. Radium*, **9,** 93.

MOMIN, A. U., 1953, *Proc. Ind. Acad. Sci.*, **37A,** 254.

MOTT, N. F., and LITTLETON, M. J., 1938, *Trans. Farad. Soc.*, **34,** 485.

PARODI, M., 1937, *C. R. Acad. Sci., Paris*, **204,** 1111, 1636, **205,** 906, 1224; 1938, *Ibid.*, **206,** 1717.

PAULING, L., 1928, *Z. Kristallogr.*, **67,** 377.

STEPHENS, R. E., and MALIKSON, I. H., 1952, *J. Res. Nat. Bur. Stand.*, 49(4), Paper 2360.

STRONG, J., 1931, *Phys. Rev.*, **37,** 72; **38,** 1565.

TOLKSDORF, S., 1928, *Z. Phys. Chem.*, **132,** 161.

WILLMOTT, J. C., 1950, *Proc. Phys. Soc.* A, **63,** 254.

WU, TA-YOU, 1946, *Vibrational Spectra and Structure of Polyatomic Molecules* (Ann Arbor, Mich. : J. W. Edwards).

The Power Radiated by an Electromechanical Wave Source

By H. PURSEY
National Physical Laboratory, Teddington, Middlesex

MS. received 14th September 1955, and in final form 4th November 1955

Abstract. In a recent paper by Evison, the author describes an electromechanical wave source for use in geophysical investigations, and presents an analysis of its behaviour. Several errors in this analysis have since been noticed, and in particular the formula for the radiation impedance of the source has now been shown to be incorrect. In the present paper a new analysis is given, based on more accurate radiation formulae which have recently been published, and taking account of the mutual admittance of the elements of the source. The results are shown to be in agreement with an estimate of radiated power made by Evison on the basis of measurements of the received signal at points distant from the source.

The present paper shows that much of the power from Evison's source is radiated in the surface wave, and a method of reducing this loss is described.

§ 1. Introduction

The use of an electromechanical transducer as a source of vibrations for geophysical investigations has been described in a recent paper (Evison 1951) in which the author analyses the behaviour of a Fessenden oscillator radiating square-enveloped pulses of energy into a medium composed of chalk. Similar radiation problems were encountered in an investigation of propagation in coal mines (Bradfield 1947), where use was made of a tentative formula for the radiation from an oscillating sphere immersed in an infinite solid. Doubts as to the validity of this formula (which was also employed by Evison) led to a re-examination of the general problem by the present writer, and the results of this work have been published in three recent papers (Pursey 1953, Miller and Pursey 1954, 1955).

The Fessenden oscillator described by Evison is coupled to the ground by three rods, situated at the vertices of an equilateral triangle. Evison assumes that the area enclosed by the circumcircle of this triangle vibrates as a rigid piston, and it is fortuitous that with this assumption the tentative formula yields a value for the radiation impedance which is in fair agreement with the true value, obtained from the new formulae by regarding the three rods as separate, small radiating elements having both self and mutual impedance components. Further errors arise in Evison's work from the use of a formula which was intended to apply to the case of a dipole source, and from neglect of the power radiated in the surface wave. To take account of the medium being semi-infinite instead of infinite, the impedance value given by the original formula has been halved, a procedure which ignores the stresses produced by the dipole source across the equatorial plane. However, it may be shown that the discrepancy between the values given by this method and by the rigorous analysis for the semi-infinite case is only of order two to one, and therefore not of great significance in comparison with probable experimental errors. The power radiated in the surface wave is a more serious source of error, since it comprises about two thirds of the total power, and may exceed this for certain configurations of the radiating elements.

In the present paper the application of the new treatment to Evison's radiator is described, and in particular the radiated power calculated by considering the input to the oscillator and the parameters of its electrical and mechanical components is shown to be in good agreement with the power estimated from measurements made by Evison of the field at a large distance from the source.

It is to be understood throughout the present paper that values of time-dependent quantities, such as r.m.s. currents and stresses, and radiated power, apply during transmission of a pulse, and are not averaged over the total recurrence period.

§ 2. CALCULATION OF THE RADIATION IMPEDANCE

The radiator has the following associated parameters: a = radius of coupling rods = 3·18 cm, s_{ij} = distance between centres of the ith and jth rods = 34·6 cm, f = frequency of vibration = 600 c/s, V_C = velocity of compressional waves in chalk = $2{\cdot}14\times10^5\,\mathrm{cm\,sec^{-1}}$, ρ = density of chalk = $1{\cdot}80\,\mathrm{g\,cm^{-3}}$, σ = Poisson's ratio for chalk = 0·25.

Now the self-admittance of each element is given by equation (30) of Miller and Pursey (1955);

$$Y_R = -\frac{2i\omega\mu^2}{\pi c_{44}}\int_0^\infty \frac{(\zeta^2-1)^{1/2}\{J_1(\zeta k_1 a)\}^2 d\zeta}{\zeta k_1 a^2 F_0(\zeta)}, \qquad \ldots\ldots(1)$$

where $\omega = 2\pi f$, $k_1 = 2\pi f/V_C$, $\mu^2 = 2(1-\sigma)/(1-2\sigma)$, $c_{44} = \rho V_C^2/\mu^2$, and

$$F_0(\zeta) = (2\zeta^2-\mu^2)^2 - 4\zeta^2(\zeta^2-1)^{1/2}(\zeta^2-\mu^2)^{1/2}.$$

The integral may be approximated asymptotically for small values of $k_1 a$. Let

$$\Phi(k_1 a) = \int_0^\infty \frac{(\zeta^2-1)^{1/2}\{J_1(\zeta k_1 a)\}^2 d\zeta}{\zeta k_1 a F_0(\zeta)}. \qquad \ldots\ldots(2)$$

Then $\Phi(k_1 a) = \Phi_0{}^\lambda + \Phi_\lambda{}^\infty$ where

$$\Phi_0{}^\lambda = \int_0^\lambda \frac{(\zeta^2-1)^{1/2}\{J_1(\zeta k_1 a)\}^2 d\zeta}{\zeta k_1 a F_0(\zeta)}, \qquad \Phi_\lambda{}^\infty = \int_\lambda^\infty \frac{(\zeta^2-1)^{1/2}\{J_1(\zeta k_1 a)\}^2 d\zeta}{\zeta k_1 a F_0(\zeta)},$$

and λ is a large, positive number, independent of a. To evaluate $\Phi_0{}^\lambda$, we put $J_1(\zeta k_1 a) \simeq \frac{1}{2}\zeta k_1 a$, and obtain

$$\Phi_0{}^\lambda \simeq \frac{k_1 a}{4}\int_0^\lambda \frac{\zeta(\zeta^2-1)^{1/2} d\zeta}{F_0(\zeta)}.$$

The value of the imaginary part of the above integral is independent of λ, being due only to the portion of the integral between 0 and μ, and the pole at $\zeta = p\ (\mu < p < \lambda)$. The value of the real part of the integral is a function of λ, and is of order a, so that for small a it may be neglected in comparison with the contribution of order unity from $\Phi_\lambda{}^\infty$. Numerical quadrature gives $\mathscr{I}(\Phi_0{}^\lambda) = 0{\cdot}1344 k_1 a$.

The term $\Phi_\lambda{}^\infty$ can be evaluated by expanding the factor $(\zeta^2-1)^{1/2}/F_0(\zeta)$ in inverse powers of ζ and integrating term by term. We thus find for the dominant term (cf. Miller and Pursey 1954, formula (132))

$$\Phi_\lambda{}^\infty \sim -\frac{1}{4k_1a}\left[\frac{2}{3}k_1a\left\{\frac{2}{\pi}+\mathrm{J}_0\mathrm{J}_1-\lambda k_1a(\mathrm{J}_0{}^2+\mathrm{J}_1{}^2)\right\}+\frac{\mathrm{J}_1{}^2}{3\lambda}\right],$$

where the argument of the Bessel functions is λk_1a. Hence, for small a, $\Phi_\lambda{}^\infty \sim -1/3\pi$. We have therefore shown that $\Phi(k_1a) \sim -1/3\pi + 0{\cdot}1344ik_1a$, and it follows from equation (1) that

$$Y_\mathrm{R} \sim \frac{2\omega k_1\mu^2}{\pi c_{44}}\left(0{\cdot}1344+\frac{i}{3\pi k_1a}\right)$$
$$\sim \frac{2\pi\mu^4f^2}{\rho V_\mathrm{C}{}^3}\left(0{\cdot}5374+\frac{4i}{3\pi k_1a}\right). \qquad \ldots\ldots(3)$$

The mutual impedance of pairs of elements is obtained from the formula

$$Y_{ij} \sim \frac{2\pi\mu^4f^2}{\rho V_\mathrm{C}{}^3}\,I(k_1s_{ij},\mu), \qquad \ldots\ldots(4)$$

where $I(x,\mu)$ is the function defined in equation (45) and tabulated in table 1 of Miller and Pursey (1955). For the present case, we find from the values listed above that $k_1a = 5{\cdot}61\times10^{-2}$ and $k_1s_{ij} = 0{\cdot}612$. It follows from equations (3) and (4) that

$$Y_\mathrm{R} = \frac{2\pi\mu^4f^2}{\rho V_\mathrm{C}{}^3}(0{\cdot}5374+7{\cdot}565i), \qquad Y_{ij} = \frac{2\pi\mu^4f^2}{\rho V_\mathrm{C}{}^3}(0{\cdot}4035-0{\cdot}0772i). \quad \ldots\ldots(5)$$

The total radiation impedance of the radiator is given by

$$Z = 3/(Y_\mathrm{R}+2Y_{ij})$$
$$= \frac{\rho V_\mathrm{C}{}^3}{2\pi\mu^4f^2(0{\cdot}4481+2{\cdot}470i)}$$
$$= (6{\cdot}106-33{\cdot}65i)\times10^7. \qquad \ldots\ldots(6)$$

§ 3. THE MECHANICAL CIRCUIT OF THE RADIATOR

Figures 1 and 2 show the mechanical and equivalent electrical networks of the radiator, using the direct analogy in which mechanical velocity is represented by electric current and mechanical force by electrical voltage.

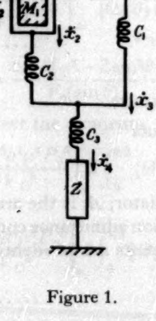


Figure 1.

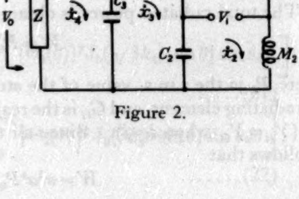


Figure 2.

The physical quantities have the following values:

M_1 = mass of magnet $=272\,\text{kg}$,
M_2 = mass of copper cylinder $=2{\cdot}72\,\text{kg}$,
C_1 = compliance of mounting rods $=4{\cdot}4\times10^{-13}\,\text{cm dyn}^{-1}$
C_2 = compliance of cylinder and clamping rod $=31{\cdot}0\times10^{-13}\,\text{cm dyn}^{-1}$,
C_3 = compliance of coupling system $=0{\cdot}9\times10^{-13}\,\text{cm dyn}^{-1}$,
V_1 = force between magnet and cylinder
V_0 = force transmitted to coupling rods
$\dot{x}_1$, etc. are displacement velocities at the points indicated in figure 1.

The mesh equations of the equivalent circuit are

$$\left.\begin{aligned} V_1=i\omega M_1\dot{x}_1+\frac{\dot{x}_1-\dot{x}_3}{i\omega C_1}&=-i\omega M_2\dot{x}_2-\frac{\dot{x}_2-\dot{x}_3}{i\omega C_2}\\ \dot{x}_3\left(\frac{1}{i\omega C_1}+\frac{1}{i\omega C_2}+\frac{1}{i\omega C_3}\right)&=\frac{\dot{x}_1}{i\omega C_1}+\frac{\dot{x}_2}{i\omega C_2}+\frac{\dot{x}_3}{i\omega C_3}\\ \dot{x}_4\left(\frac{1}{i\omega C_3}+Z\right)&=\frac{\dot{x}_3}{i\omega C_3}\\ V_0&=\dot{x}_4 Z. \end{aligned}\right\}\quad\ldots\ldots(7)$$

Solving these equations for V_0 in terms of V_1, we obtain

$$V_1 Z\left\{\frac{\omega^2(M_1C_1-M_2C_2)}{(1-\omega^2M_1C_1)(1-\omega^2M_2C_2)}\right\}$$
$$=V_0\left\{\frac{i\omega M_1}{1-\omega^2M_1C_1}+\frac{i\omega M_2}{1-\omega^2M_2C_2}+Z\left(1-\frac{\omega^2M_1C_3}{1-\omega^2M_1C_1}-\frac{\omega^2M_2C_3}{1-\omega^2M_2C_2}\right)\right\}.\quad\ldots\ldots(8)$$

Inserting numerical values of the parameters in the above equation gives

$$|V_0/V_1|=0{\cdot}448.\quad\ldots\ldots(9)$$

The value of V_1 may be obtained from the expression $V_1=10^{-1}BLI$, where B is the magnetic flux through the copper cylinder, L is the circumferential length of the cylinder, and I is the current in amps flowing in the cylinder. The figures given by Evison are $B=7{\cdot}25\times10^{-3}$ gauss, $L=64$ cm, and the current in the cylinder may be deduced from the information that 7·3 amps r.m.s. flow in the a.c. coil of 520 turns surrounding the cylinder, giving a current of 3800 amps r.m.s. in the cylinder. We find that $V_1=1{\cdot}76\times10^8$ dyn, and therefore $V_0=7{\cdot}89\times10^7$ dyn. Hence, for the normal stress under the coupling rods we obtain

$$P_0=V_0/3\pi a^2=8{\cdot}33\times10^5\,\text{dyn cm}^{-2}.\quad\ldots\ldots(10)$$

§ 4. CALCULATION OF RADIATED POWER

The total radiated power is obtained from the formula

$$W=P_0{}^2\sum_{i,j}A_iA_jG_{ij},\quad\ldots\ldots(11)$$

where P_0 is the r.m.s. value of the stress under the radiator, A_i is the area of the ith radiating element, and G_{ij} is the real part of the radiation admittance component Y_{ij} ($Y_{ij}=Y_R$ when $i=j$). Since all the radiating elements are of identical size, it follows that

$$W=\pi^2a^4P_0{}^2(3G_R+6G_{ij(i\neq j)}).\quad\ldots\ldots(12)$$

From equation (5) we obtain

$$3G_{\mathrm{R}}+6G_{ij}=\frac{2\pi\mu^4 f^2}{\rho V_{\mathrm{C}}^3}(4\cdot 033)=4\cdot 698\times 10^{-9}. \qquad \ldots\ldots(13)$$

Hence, from equations (10), (12) and (13),

$$W=3\cdot 25\times 10^6\,\mathrm{erg\,sec^{-1}}=0\cdot 325\ \mathrm{watt}. \qquad \ldots\ldots(14)$$

To calculate the power radiated in the surface wave, we use equation (41) of Miller and Pursey (1955):

$$W_{\mathrm{Su}}=\frac{8\pi^3 f^2 a^4 \mu^2 P_0^{\,2}}{\rho V_{\mathrm{C}}^3}\int_0^\infty X(k_1 z)\,d(k_1 z)\sum_{l=1}^{n}\sum_{m=1}^{n}\mathrm{J}_0\left\{2pk_1 b\sin\frac{\pi(l-m)}{n}\right\}, \quad \ldots\ldots(15)$$

where p is the affix of the pole on the real axis in the integral of equation (1), equal to 1·8839 for media of Poisson's ratio $\frac{1}{4}$, b is the radius of the circle on which the radiating elements are situated, n is the number of radiating elements, and

$$X(k_1 z)=1\cdot 5383\exp(-1\cdot 4820k_1 z)-3\cdot 1192\exp(-2\cdot 3376k_1 z)$$
$$+1\cdot 8131\exp(-3\cdot 1932k_1 z).$$

The additional factor of 2 in equation (15) takes account of the fact that r.m.s. values of alternating quantities are used throughout the present paper. The integral in equation (15) is readily evaluated, and we obtain

$$W_{\mathrm{Su}}=\frac{6\cdot 512\pi^3 f^2 a^4 P_0^{\,2}}{\rho V_{\mathrm{C}}^3}\{3+6\mathrm{J}_0(3\cdot 263k_1 b)\}. \qquad \ldots\ldots(16)$$

Substituting numerical values for the parameters in the above expression gives for the power radiated in the surface wave

$$W_{\mathrm{Su}}=0\cdot 208\ \mathrm{watt}. \qquad \ldots\ldots(17)$$

Subtracting the power in the surface wave from the total radiated power gives for the power radiated in the shear and compressional waves the value of 0·117 watt. This is in good agreement with the figure of 0·071 watt, estimated by Evison from measurements of the received signal at a distance from the source.

Finally, to obtain the separate powers in the compressional and shear waves, we use equations (39) and (40) of Miller and Pursey (1955):

$$W_{\mathrm{C}}=\frac{2\pi^3 f^2 a^4 \mu^4 P_0^{\,2}}{\rho V_{\mathrm{C}}^3}\int_0^{\pi/2}\{\Theta_1(\theta)\}^2\sum_{l=1}^{n}\sum_{m=1}^{n}\mathrm{J}_0\left\{2\mathrm{k}_1 b\sin\theta\sin\frac{\pi(l-m)}{n}\right\}\sin\theta\,d\theta, \quad \ldots(18)$$

$$W_{\mathrm{Sh}}=\frac{2\pi^3 f^2 a^4 \mu^9 P_0^{\,2}}{\rho V_{\mathrm{C}}^3}\int_0^{\pi/2}|\Theta_2(\theta)|^2\sum_{l=1}^{n}\sum_{m=1}^{n}\mathrm{J}_0\left\{2\mu k_1 b\sin\theta\sin\frac{\pi(l-m)}{n}\right\}\sin\theta\,d\theta, \ldots(19)$$

where

$$\Theta_1(\theta)=\frac{\cos\theta(\mu^2-2\sin^2\theta)}{F_0(\sin\theta)}, \qquad \Theta_2(\theta)=\frac{\sin 2\theta(\mu^2\sin^2\theta-1)^{1/2}}{F_0(\mu\sin\theta)}. \qquad \ldots\ldots(20)$$

Summing over the elements, we obtain

$$W_{\mathrm{C}}=\frac{2\pi^3 f^2 a^4 \mu^4 P_0^{\,2}}{\rho V_{\mathrm{C}}^3}\left[3\int_0^{\pi/2}\{\Theta_1(\theta)\}^2\sin\theta\,d\theta+6\int_0^{\pi/2}\{\Theta_1(\theta)\}^2\mathrm{J}_0(\sqrt{3}k_1 b\sin\theta)\sin\theta\,d\theta\right], \qquad \ldots\ldots(21)$$

$$W_{\mathrm{Sh}}=\frac{2\pi^3 f^2 a^4 \mu^9 P_0^{\,2}}{\rho V_{\mathrm{C}}^3}\left[3\int_0^{\pi/2}|\Theta_2(\theta)|^2\sin\theta\,d\theta+6\int_0^{\pi/2}|\Theta_2(\theta)|^2\mathrm{J}_0(3k_1 b\sin\theta)\sin\theta\,d\theta\right]. \qquad \ldots\ldots(22)$$

The integrals in the above expressions have been evaluated numerically for the case of $\mu=\sqrt{3}$ and $k_1b=0{\cdot}3534$, and we find that

$$\int_0^{\pi/2}\{\Theta_1(\theta)\}^2\sin\theta\,d\theta=0{\cdot}03700,$$

$$\int_0^{\pi/2}|\Theta_2(\theta)|^2\sin\theta\,d\theta=0{\cdot}008878,$$

$$\int_0^{\pi/2}\{\Theta_1(\theta)\}^2\mathrm{J}_0(0{\cdot}6121\sin\theta)\sin\theta\,d\theta=0{\cdot}03557,$$

$$\int_0^{\pi/2}|\Theta_2(\theta)|^2\mathrm{J}_0(0{\cdot}0602\sin\theta)\sin\theta\,d\theta=0{\cdot}007486.$$

Substitution of these values in the expressions (21) and (22) gives

$$W_{\mathrm{C}}=0{\cdot}026\text{ watt},\qquad W_{\mathrm{Sh}}=0{\cdot}090\text{ watt}.\qquad\ldots\ldots(23)$$

We therefore obtain for the percentage of power in the three waves:

Compressional wave	8%
Shear wave	28%
Surface wave	64%.

It may be shown (see Miller and Pursey 1955) that by so spacing the radiators as to give minimum value to the factor $\{3+6\mathrm{J}_0(3{\cdot}263k_1b)\}$ in equation (16) the percentage of power radiated in the surface wave will be reduced to 26%, and accordingly the efficiency of the radiator as a source of compressional and shear waves will be substantially increased. The corresponding value of k_1b is 1·1743, so that the distance between elements is an appreciable fraction of the compressional wavelength; the polar radiation patterns will therefore no longer approximate to those of a point source, and in particular there will now be some variation of field strength with azimuth.

ACKNOWLEDGMENTS

The author wishes to acknowledge the contribution of Mrs. O. E. Taylor, who computed the numerical integrals involved in expressions for radiated power; acknowledgment is also due to Mr. G. Bradfield and Mr. G. F. Miller for discussion and criticism of the paper.

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

REFERENCES

BRADFIELD, G., 1947, *Telecommunications Research Establishment Memorandum*, Z/M120/GB.
EVISON, F. F., 1951, *Proc. Phys. Soc.* B, **64**, 311.
MILLER, G. F., and PURSEY, H., 1954, *Proc. Roy. Soc.* A, **223**, 521; 1955, *Ibid.*, **233**, 55.
PURSEY, H., 1953, *Brit. J. Appl. Phys.*, **4**, 12.

Ferromagnetic Domain Nucleation in Silicon Iron

By L. F. BATES and D. H. MARTIN†
Physics Department, University of Nottingham

MS. received 19*th August* 1955 *and in final form* 31*st October* 1955

Abstract. Powder deposit patterns obtained by the authors on single crystals of 3% silicon iron are consistent with a description of the processes of domain phase creation in terms of nucleation at inclusions in the surfaces. Domain nucleation is discussed in relation to various magnetic properties.

---

§ 1. INTRODUCTION

SINCE our earlier communication on powder deposit studies of domain nucleation (Bates and Martin 1953, denoted by I) interest in the subject has increased (Martin 1953, Goodenough 1954, Menyuk and Goodenough 1955). Further information on this important topic and also a fuller description of the domain processes following upon nucleation are given below. The powder deposits, produced by the technique described, for example, by Bates and Neale (1950), were obtained on a single crystal disc of 3% Si–Fe, 7·5 mm in diameter and 0·13 mm in depth, whose main faces were cut parallel to the (001) plane.

The work of Néel (1944) and of Lawton and Stewart (1948) showed that domain structures of crystals are, in general, composed of finite numbers of domain phases each characterized by the common direction of magnetization of the domains comprising the phase. In iron, for example, the demagnetized state is normally a six-phase state, while the saturated condition is single-phase. In intermediate fields, two, three and four-phase 'modes of magnetization' may be stable. Whenever a domain structure changes from one mode to another characterized by a larger number of phases, the formation of the domains constituting the new phases must in general be initiated by nucleation processes of some kind. Two such changes of mode, viz. one-phase to four-phase and one-phase to two-phase, were considered in our earlier communications. The first case, and a further new one—namely two-phase to four-phase—are more fully examined here.

§ 2. EXPERIMENTAL RESULTS

2.1. *Single-phase to Four-phase Transitions*

Before being subjected to a progressively decreasing applied field, during which process the domain nucleation occurred, the specimen was saturated to achieve the single-phase state in the [100] direction, which was an easy direction of magnetization in the plane of the disc. As the applied field was increased to produce saturation, the domain walls in the surface moved to bring about annihilation of the domains magnetized unfavourably with respect to the effective field, until, in a relatively low field, the domain structure consisted only of 'spike' domains based on small pits in the surface and magnetized perpendicularly to the main phase in which they were set. With further increase in field these spikes diminished in length, and for an effective field of about 150–200 oersteds, as measured by a magnetic potentiometer, the 'spikes' were too short to be distinguished from the clouds of colloid particles which collected at the surface pits or inclusions. They were then shorter than the diameters of the parent imperfections. The process was very similar to that shown in figure 1 (Plates I, II) but in reverse. The effective field was increased to some 800 oersteds and demagnetization by reduction of the field was then started.

---

† Now at Queen Mary College, University of London.

The progressive break-up of the single domain of the saturated specimen as the applied field was reduced, is shown in the composite photographs, figure 1(*a*) to (*f*), of the powder patterns on a chosen small region of the specimen's (001) surface as the applied field was decreased from (*a*) to (*f*). For figure 1(*a*) the effective field which ran parallel to the edge of the picture was reduced to about 200 oersteds. Small clouds of colloid particles are seen at numerous small pits in the surface and show them up clearly, but there is no sign of domain walls. For figure 1(*b*) the effective field was reduced to about 150 oersteds. At many of the surface pits 'spike' domains are evident in the photograph; some whose bases are at least as small as 5 microns can be seen, while the large spike in the photograph is based on a surface pit, of diameter about 50 microns, of which part is seen at the edge of the field of view in figures 1 (*b*)–(*f*). The spike domains are magnetized transversely to the surrounding phase, i.e. in the [010] direction. This is indicated not only by their direction, the [$\bar{1}$10], which is such that the normal component of magnetization is nearly continuous across the spike walls, but also, in the larger spikes, by the colloid striations which are always perpendicular to the underlying magnetization. These spike domains will now be called *transverse domains*. (The surface was unfortunately stained by the colloid in one region to be noticed at the right edges of the photographs; this stain should not be confused with surface pits.)

For figure 1 (*c*) the effective field was reduced to approximately 70 oersteds, and new features appear. At several points domains can be seen whose walls lie in the up and down direction, i.e. the [100] direction, and the one associated with the above-mentioned large transverse domain is especially clear. These domains are magnetized in the *opposite* sense to the main domain phase. We will now call them *reverse domains*, and their origins will be discussed later.

Further decreases in the applied field brought about the changes illustrated in figures 1 (*d*), (*e*) and (*f*). Figures 1 (*d*) and (*e*) show how both the transverse and reverse domains grow, while others appear, although some eventually decrease in size. Finally, in figure 1 (*f*) are seen large domains formed by the coalescing of several of the transverse or reverse domains, and these are bounded by straight walls. The specimen is now in a four-phase state. No field values are quoted for these composite photographs, because the small size of the specimen made effective field measurements difficult, and in any case the effective fields were probably not sufficiently uniform for the measured values to be reliably attached in those low field conditions to the structures observed in any particular region. The effective field for figure 1(*f*) may well have been negative, because of the demagnetization field, even though the applied field was itself small and positive; the same statement may apply to figure 1 (*e*). The mean intensity of magnetization in the surface layer, however, may be estimated from the relative volumes occupied by the several domain phases, as revealed by the powder patterns; it is about

$0{\cdot}7 \times I_s$ in figure 1(*e*) and $0{\cdot}5 \times I_s$ in figure 1(*f*), where I_s is the saturation magnetization. The patterns therefore correspond to the steep portion of the (de)magnetization curve.

The complicated barbs attached to the reverse domain nuclei should not be confused with the tree-type structures found on 90° and 180° walls in surfaces which are not exactly parallel to the (001) plane. The surface of the specimen was within $\frac{1}{3}°$ of the (001) plane, as is manifest by the straight, clear boundaries of the large domains in figure 1 (*f*). These complicated forms of reverse domain nuclei are features common to all the cases studied.

A general description of the process of domain nucleation as observed in the specimen surface will now be given with reference to the photographs. The single domain of the saturated crystal first breaks down at certain of the microscopic pits in the surface, with the formation there of short spike domains based upon the pits and magnetized transversely to the direction of the magnetization of the surrounding phase. These domains are first observed when their lengths are about the same as their widths, the latter being approximately equal to the diameters of the parent pits. The effective field at this stage is between 150 and 200 oersteds, and there is no apparent systematic dependence of this threshold field upon the size of the imperfections. Figure 2(*a*) (Plate III) shows a further example of a saturated portion of the crystal's surface with the surface defects clearly revealed by powder deposits, and figure 2(*b*) shows the appearance of spike transverse domains at some of the pits when the field is reduced to about 150 oersteds. Figure 3 (Plate III) shows numerous spike domains at a badly pitted part of the surface.

The transverse domains expand markedly, especially in length, when the applied field is further decreased, and not until the effective field is about 75 oersteds do the first reverse domains appear. These domains are generally attached to one or other of the transverse domains, a point to be more fully discussed below. Figure 4 (*a*) (Plate III) is a further illustration of transverse domains and figure 4(*b*) shows how, for a small reduction in field, a marked rearrangement takes place with the appearance of reverse domains with walls in the [100] direction. Figure 5 (Plate III) shows the close association of reverse and transverse domains. Further reduction in the applied field brings about marked expansion of some of the transverse and reverse domain nuclei. This process takes place at the knee of the (de)magnetization curve, where the crystal adopts a four-phase, mode I, structure. The details of this part of the process are complicated, with some nuclei at first expanding and then later contracting, while by some process of selection certain of the nuclei of transverse and of reverse magnetization grow or combine to form the large domains of the four-phase state.

The complete details of the whole process, from nucleation to the final demagnetized state, are never repeated in a subsequent run, not even when a few large pits of diameter about 50 microns are present. This makes detailed study of the exact origins of reverse domain nuclei extremely difficult, and the situation is further aggravated by the fact that, although a slowly changing domain structure can be followed with a suitable colloid, a rapid change in the structure generally leaves the patterns behind. The old patterns take time to disintegrate, about 30 seconds, and the new patterns take time to form, and so a change in structure is not immediately revealed. In the early stages of nucleation the forms of the reverse domains are very sensitive to field, yet very slow changes in the applied field are precluded because of the tendency of the patterns to become permanent if stationary for more than a minute or two, and also because of the eventual staining of the polished surface by the powder suspension. The exact process of nucleation is therefore not easily observed and is even less readily recorded.

2.2. *Two-phase to Four-phase Transitions*

The same crystal as before was used, but it was now magnetized in the [110] direction which was in the disc surface between the [010] and [100] easy directions. When an effective field acts along the [110] axis of an iron crystal a two-phase, mode III, structure is stable. This is made up of domains, whose lengths lie along the $[\bar{1}10]$ direction, and which are alternately magnetized along directions making equal angles, zero in small fields, with the [010] and [100] directions, respectively; (cf. the Néel cut, Bates and Neale 1950). Figure 6(*a*) (Plate IV) shows a fraction of the domain structure of the disc when subjected to a small field in the [110] direction. The colloid striations indicate the directions of magnetization of the three major domains, which are illustrated diagrammatically in figure 6(*h*). In order that the specimen may enter a four-phase mode I state when the applied field is reduced, domains magnetized in the $[\bar{1}00]$ and $[0\bar{1}0]$ directions have to appear.

In the central and lower domains of figure 6(*a*) are seen several small spike domains at pits in the surface; one such is encircled in the picture. They are each magnetized perpendicularly to their parent domains and in the senses favoured by the applied field, i.e. in the [100] or [010] directions. As the applied field was progressively reduced the photographs figure 6(*b*) to (*g*) were taken. For figure 6(*b*) the spike domains have grown, but no domain magnetized in the $[\bar{1}00]$ or the $[0\bar{1}0]$ is yet present. In figure 6(*c*), however, two such domains, magnetized in the $[\bar{1}00]$ direction, are seen attached to the largest spike in the central main domain, with their walls running parallel to the [110] direction. In figure 6(*d*) there appears a domain magnetized in the $[0\bar{1}0]$ direction, again attached to the largest spike and with its walls running at 45° to the horizontal. This domain expands considerably to give figure 6(*e*). Another domain magnetized in the $[0\bar{1}0]$ direction appears here, this time in the lower main domain, with its walls running in the vertical direction, and again apparently associated with one of the original spike domains. The specimen has now commenced to demagnetize itself, being at the knee of the (de)magnetization curve, and subsequently the domains of the new $[\bar{1}00]$ and $[0\bar{1}0]$ phases grow rapidly to form a four-phase state like that shown in figure 6(*g*). The general process of demagnetization, as revealed by the surface domains structures, therefore resembles that described in § 2.1. In particular, there occurs a sequence of creation and growth of spike domains *followed* by the appearance and growth of reverse domains.

§ 3. Discussion of Results

It is well known (cf. Kittel 1949) that there is a theoretical lower limit, given by $2K/I_s$, for the critical reverse field required to initiate nucleation in a perfect crystal by a rotation-of-magnetization mechanism; K is the appropriate anisotropy constant and I_s is the spontaneous magnetization. For iron $2K/I_s$ is about 500 oersteds, yet the coercivity is seldom greater than an oersted or so. It is therefore apparent that nuclei are not created in an ordinary crystal under the action of the applied field alone.

In a saturated specimen there must exist close to any non-magnetic inclusion additional fields of at least the order of I_s, which exceed $2K/I_s$ in the case of iron. Nucleation by rotation close to inclusions is therefore possible. This is the process which occurs in the surfaces of our crystals where domain nuclei were first observed at surface pits, i.e. non-magnetic inclusions. Nucleation in a perfect crystal under the action of a uniform applied field would be irreversible; the nucleus, once formed, would grow rapidly. In the present mechanism of nucleation by rotation at inclusions, however, the promoting field has only a restricted range, and, moreover, the onset of rotation serves to decrease the promoting field. It seems probable, therefore, that the rotation, which finally produces a recognizable, short, transversely magnetized spike domain, occurs over a large range of applied field and only over a volume comparable with that of the inclusion.

The surface pits in the present work, and in I, provided numerous *transverse* domain nuclei in the manner just discussed. No observations were made, however, of a similar direct creation of *reverse* domains at surface pits. Rotation at an inclusion would not be expected to produce domains oppositely magnetized to the surrounding phase, except in uniaxial materials, since rotation beyond the transverse direction would generally serve only to increase the magnetostatic energy. The delayed appearance of the reverse domains, and their initial attachment to transverse domains in most cases, suggest that they may be produced as the result of encounters between inclusions and the moving 90° walls of transverse domains. A model of this process was suggested in I and receives support from many of the details observed during the present work, some of which are to be found in our figures, and also from recent observations by Martin (1954, see Bates 1954), of encounters between 90° walls and large surface pits. These latter observations showed that processes of the types illustrated in figure 7 actually occur, each producing a small reversely magnetized domain which in favourable circumstances could grow to form a reverse domain nucleus. Thus, in figure 7 (*a*) the 90° wall WW moves under the action of a changing applied field from its initial position in figure 7 (*a*) (i) to pass through the inclusion, shown as a hatched square, to produce the reverse domain nucleus R of figure 7 (*a*) (iv). In figure 7 (*b*) the domain system on the left of figure 7 (*b*) (i) moves across from left to right to strike the inclusion, forming in figure 7 (*b*) (iii) the reverse domain R.

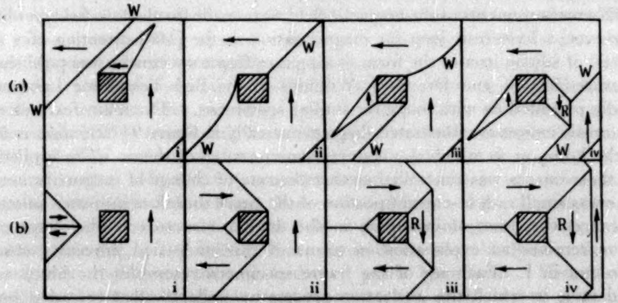


Figure 7. Illustrations of two types of interaction between moving 90° walls and inclusions. Reverse domain nuclei are labelled R.

Possible mechanisms, based on these processes, which would serve to produce reverse domains during encounters between the moving 90° walls of transverse domains and inclusions are illustrated in figure 8, and domain configurations

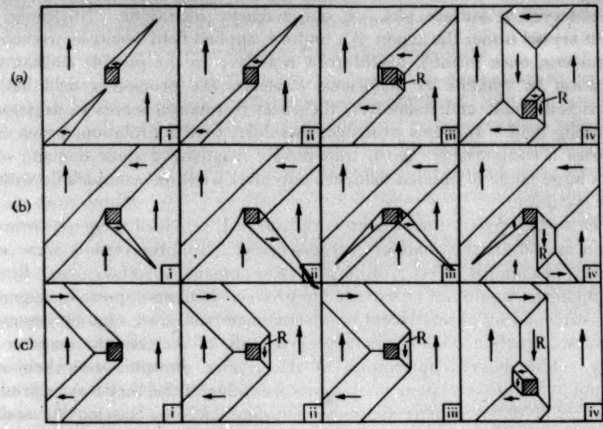


Figure 8. Illustration of three processes, (*a*), (*b*) and (*c*), of reverse domain creation as results of encounters between expanding transverse domains and inclusions.

like the end-products of these mechanisms can be found in several places in our photographs. Most of the reverse domains in our figures show strong circumstantial evidence of having been initially created as results of such encounters between 90° walls and surface pits; that associated with the large spike domain in figure 1 (*b*) and (*c*) may be especially mentioned. Connections between a few of the reverse domains and transverse structures are not obvious. However, this fact may be explained by the rapid movements of favourable nuclei when once formed, or the possibility of transverse structures being created below the surface or in the opposite face and growing until they met the surface under observation.

No measurements of the magnetic field were made for the low-field conditions. However, a hysteresis loop for magnetization in the [100] direction of a single crystal of silicon iron in the form of a picture frame specimen, was published by Stewart (1951); and Dr. H. J. Williams of the Bell Telephone Laboratories kindly provided us with loops for similar specimens. The main features of the hysteresis curves are illustrated diagrammatically in figure 9; attention is drawn to the hump at b which is a feature common to all of them. The applied field for these curves was controlled so that the rate of change of magnetization with time was small. A re-entrant portion of the curve therefore indicates an attempt to cancel the onset of instability in the domain structure. The shape of the curve receives an explanation in terms of the nucleation processes observed here and in I. Each leg of the frame specimens resembles the block crystal used in I, in which the nucleation process is similar to that recorded here in §2.1 except that, in the former case, a large reverse domain, which expanded along the whole length of the crystal, tended to form by the fusion of several reverse nuclei. Such a large reverse domain would provide the single 180° wall which has been observed (Williams and Shockley 1949) to move across the leg of a picture frame specimen, and whose movement corresponds to the portion cd of figure 9. The region ab would therefore correspond to the nucleation and growth of transverse and, later, reverse domains as recorded in our photographs. The pronounced re-entrant portion bc indicates the onset of instability in one or other of the reverse domains with the consequent formation of the plane 180° wall.

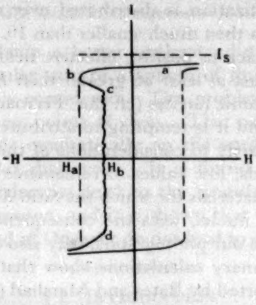


Figure 9. Diagram of re-entrant demagnetization curve for single crystals of silicon iron. For Williams crystal, $H_c = 0{\cdot}038$ Oe, $H_b = 0{\cdot}005$ Oe. For Stewart crystal, $H_c = 0{\cdot}050$ Oe, $H_b = 0{\cdot}035$ Oe. (H_c = coercivity.)

Döring (1938) studied the stability of an ellipsoidal reverse domain and concluded that the length of such a domain of fixed equatorial diameter a approached infinity asymptotically as the reverse field approached the value $H_n = \frac{3}{4}\pi\gamma/I_s a$ where γ is the energy per square centimetre of the domain wall. If one makes a formal substitution in this equation, one obtains $a = 0{\cdot}05$ cm and $a = 0{\cdot}1$ cm for the Williams and the Stewart curves respectively, using $H_n = H_c - H_b$ since, in expanding, the nucleus may have to overcome the impedance to wall motion indicated by the field H_b in figure 9. In view of the differences between the nuclei observed in this work and the ellipsoidal shape considered by Döring, and the neglect of the probable merging of domains before the onset of instability, these values, representing upper limits for the diameters of the largest nuclei present, appear to be of reasonable magnitude.

In polycrystalline material the available inclusions within the grains would generally be smaller than the larger pits in the surfaces of a single crystal. Goodenough (1954) concluded, from comparisons of free energies before and after hypothetical nucleation, that grain boundaries should be the most important sources of nuclei. He did not consider actual mechanisms of nucleation, in particular those observed in I and those recorded here. Before nucleation can take place in a strain-free material the field at the point in question must at least be greater than $2K/I_s$, and fields of this magnitude are more likely to be found at non-magnetic inclusions than at ordinary grain boundaries. Nevertheless, the additional stray fields due to random orientation of the crystal grains may materially assist nucleation at inclusions and the subsequent growth of the domains, especially those at inclusions at or close to certain grain boundaries.

Transverse domains originated at the edges of the single crystal used in the present work and played a part in the demagnetization process similar to that of the transverse structures nucleating at pits within the crystal face.

The additional local fields caused by a non-magnetic inclusion are of magnitude cI_s, where the factor c may be about 10, except in materials of very low K in which the divergence of magnetization is distributed over a larger volume than that of the inclusion so that c is then much smaller than 10. If nucleation in a strain-free material is to take place in positive effective fields, i.e. those favouring the parent phase, then cI_s must at least be greater than K/I_s. This is the case for iron, but *not* the case for some ferrites (cf. the 'Ferroxdure' sample of Rathenau, Smit and Stuyts 1952), and it is tempting to attribute the rectangular hysteresis loops of certain ferrites partly to considerations of this kind. When nucleation is possible in positive fields, low values of remanence may be expected and are, in fact, found with some materials for which the ratio K/I_s^2 is low (Lee 1952).

The expansion of the nuclei, with the consequent increase in total domain wall area so very evident in our photographs, may also contribute to the magneto-caloric effect, and preliminary calculations show that this feature may explain the anomalous effects reported by Bates and Marshall (1953).

Centres of strain have been observed by Martin (1954, see Bates 1954) to provide subsidiary domains in iron similar to those associated with inclusions, and must therefore be regarded as potential nucleation centres especially in materials with a high magnetostriction constant.

In conclusion it should perhaps be noted that a nucleation process, to be distinguished from the phase transition type discussed here, will occur when the domains of a given mode multiply as a result of a decrease in the optimum domain size. In these cases the new domains must appear in the surface whose magnetic energy density is reduced by the multiplication. Such a process has been observed by the writers when a field applied to a Néel cut crystal is increased, and a similar case has been reported by Dijkstra and Martius (1953) for silicon iron under tension.

ACKNOWLEDGMENT

We are greatly indebted to the Electrical and Allied Industries Research Association for support which made this work possible, and for permission to publish this paper.

REFERENCES

BATES, L. F., 1954, *J. Inst. Met.*, **82**, 417.
BATES, L. F., and MARSHALL, G., 1953, *Rev. Mod. Phys.*, **25**, 17.
BATES, L. F., and MARTIN, D. H., 1953, *Proc. Phys. Soc.* A, **66**, 162.
BATES, L. F., and NEALE, F. E., 1950, *Proc. Phys. Soc.* A, **63**, 374.
DIJKSTRA, L. J., and MARTIUS, U. M., 1953, *Rev. Mod. Phys.*, **25**, 146.
DÖRING, W., 1938, *Z. Phys.*, **108**, 137.
GOODENOUGH, J. B., 1954, *Phys. Rev.*, **95**, 917.
KITTEL, C., 1949, *Rev. Mod. Phys.*, **21**, 575.
LAWTON, H., and STEWART, K. H., 1948, *Proc. Roy. Soc.* A, **193**, 72.
LEE, E. W., 1952, *Proc. Phys. Soc.* B, **65**, 455.
MARTIN, D. H., 1953, *Proc. Phys. Soc.* B, **66**, 712; 1954, *Thesis*, University of Nottingham.
MENYUK, N., and GOODENOUGH, J. B., 1955, *J. Appl. Phys.*, **26**, 8.
NÉEL, L., 1944, *J. Phys. Radium*, **5**, 241.
RATHENAU, G. W., SMIT, J., and STUYTS, A. L., 1952, *Z. Phys.*, **133**, 250.
STEWART, K. H., 1951, *J. Phys. Radium* **12**, 325.
WILLIAMS, H. J., and SHOCKLEY, W., 1949, *Phys. Rev.*, **75**, 178.

On the Molecular Theory of Electrostriction

By B. K. P. SCAIFE
Dublin Institute for Advanced Studies

MS. received 5*th September* 1955

Abstract. In order to obtain a better understanding of the phenomenon of electrostriction, the particular case of a sphere of fluid dielectric, subjected to a uniform external field, is considered. The problem is discussed from both macroscopic and microscopic points of view. That part of the spur of the stress tensor $3\Delta P$ which depends quadratically on the applied field is calculated for both non-polar and dipolar dielectrics. It is found that the Lorenz–Lorentz and Debye theories of dielectrics lead to the singular result that ΔP is zero. The reason for this is discussed, and it is shown that the more modern theories of Yvon and Kirkwood and of Onsager do not lead to such a result.

---

§ 1. Introduction

The word electrostriction is used in this paper to refer to the change in pressure which occurs in a dielectric fluid when subjected to the influence of an electric field. Further, attention is confined to effects which depend quadratically on the applied field.

Most discussions of this phenomenon (e.g. Abraham and Becker 1932, Bruhat and Pauthenier 1925, Guggenheim 1949) are given from a macroscopic point of view. The case usually considered is that of a parallel plate condenser immersed in a very large volume of a fluid dielectric. However, from a microscopic point of view, this case is difficult to discuss because the number of molecules within the condenser plates is a function of the applied field.

An attempt is made here to develop a microscopic theory of electrostriction. A very particular case is considered, namely that of a spherical specimen of a fluid dielectric. The reason for this choice is that it enables one to understand more easily the mechanism which is responsible for electrostriction.

It is shown in the later sections that the Lorenz–Lorentz and Debye theories (see Debye 1929) indicate that electrostriction should not occur in a spherical specimen but that more elaborate theories do predict such an effect. The reason for the inadequacy of simple dielectric theory, in this instance, is discussed.

It must be appreciated that the change in pressure due to electrostriction in fluids is extremely small even in the case of very strong applied fields. Frivold (1921) has carried out measurements of the effect in gases. The effect in liquids is more difficult to observe because of the very low compressibility.

§ 2. Macroscopic Theory

Consider a spherical specimen of a fluid dielectric subjected to a uniform external electric field E. The volume of the sphere will be denoted by V, its temperature by T and the pressure of the fluid by P. P is the mean hydrostatic pressure and is equal to one third of the spur of the stress tensor, P_{ij}, thus

$$P = \tfrac{1}{3}\sum_{i=1}^{3} P_{ii}. \qquad \ldots\ldots(1)$$

The internal energy $\overline{U}$, of the specimen satisfies the relation,

$$d\overline{U}=dQ-PdV-Ed(\gamma E), \qquad \ldots\ldots(2)$$

where dQ is the change in heat content of the system and γ is the polarizability of the specimen. It is assumed that surface effects may be ignored.

The Helmholtz free energy F is given by

$$F=U-TS \qquad \ldots\ldots(3)$$

where S is the entropy. For an isothermal process

$$dF=d\overline{U}-TdS, \qquad \ldots\ldots(4)$$

and since $dQ=TdS$, we find that

$$dF=-PdV-Ed(\gamma E). \qquad \ldots\ldots(5)$$

Choosing V, E^2 and T as independent variables, we obtain

$$d(\gamma E)=E\left(\frac{\partial\gamma}{\partial V}\right)_{E^2,T}dV+\left\{E\left(\frac{\partial\gamma}{\partial E^2}\right)_{V,T}+\frac{1}{2}\left(\frac{\gamma}{E}\right)\right\}d(E^2)+E\left(\frac{\partial\gamma}{\partial T}\right)_{V,E^2}dT. \qquad \ldots\ldots(6)$$

We now introduce the assumption that

$$\left(\frac{\partial\gamma}{\partial E^2}\right)_{V,T}=0. \qquad \ldots\ldots(7)$$

This condition implies that γ depends upon E only as a result of a change of density. For an isothermal process, the introduction of (7) into (6) leads to the equation

$$d(\gamma E)=E\left(\frac{\partial\gamma}{\partial V}\right)_{E^2,T}dV+\frac{1}{2}\left(\frac{\gamma}{E}\right)d(E^2); \qquad \ldots\ldots(8)$$

consequently (5) becomes

$$dF=-PdV-E^2\left(\frac{\partial\gamma}{\partial V}\right)_{E^2,T}dV-\tfrac{1}{2}\gamma d(E^2). \qquad \ldots\ldots(9)$$

Because dF must be an exact differential, we conclude, from (9), that

$$\left(\frac{\partial P}{\partial E^2}\right)_{V,T}=-\frac{1}{2}\left(\frac{\partial\gamma}{\partial V}\right)_{E^2,T}. \qquad \ldots\ldots(10)$$

We may integrate (10) and derive the result that ΔP, the change in pressure, is given by

$$\Delta P=-\tfrac{1}{2}E^2\left(\frac{\partial\gamma}{\partial V}\right)_{E^2,T}. \qquad \ldots\ldots(11)$$

§ 3. THE EVALUATION OF $(\partial\gamma/\partial V)_{E^2,T}$ ACCORDING TO THE THEORIES OF DEBYE AND ONSAGER

The first step is to relate γ, the polarizability of the sphere, with ϵ, the dielectric constant of the fluid. A simple calculation shows that

$$\gamma=R^3[(\epsilon-1)/(\epsilon+2)], \qquad \ldots\ldots(12)$$

were R is the radius of the sphere. It follows therefore, that

$$\left(\frac{\partial\gamma}{\partial V}\right)_{E^2,T}=\frac{3\pi}{4}\left[\frac{\epsilon-1}{\epsilon+2}\right]+\frac{3\pi}{4}\,\frac{3}{(\epsilon+2)^2}\left(\frac{\partial V}{\partial\epsilon}\right)_{E^2,T}. \qquad \ldots\ldots(13)$$

According to Debye's theory (1929) of the dielectric constant,

$$[(\epsilon-1)/(\epsilon+2)]V=(4\pi N/3)[(\beta\mu^2/3)+\alpha]. \qquad \ldots\ldots(14)$$

On the other hand, Onsager's theory (Onsager 1936) leads to the relation

$$3[(\epsilon-n^2)(2\epsilon+n^2)/\epsilon(n^2+2)^2]V=(4\pi N\beta\mu^2/3). \qquad \ldots\ldots(15)$$

In equations (14) and (15), N is the number of molecules in volume V, μ is the vacuum dipole moment of a molecule, α the combined atomic and electronic polarizability of a molecule, n the refractive index and $\beta=(kT)^{-1}$ where k is Boltzmann's constant and T is the absolute temperature.

When $\mu=0$, the Debye and Onsager theories coincide and yield the formula

$$[(n^2-1)/(n^2+2)]V=(4\pi N\alpha/3), \qquad \ldots\ldots(16)$$

which is the Lorenz–Lorentz relation.

From (14) we obtain

$$\left(\frac{\partial\epsilon}{\partial V}\right)_{\text{Debye}}=-\frac{(\epsilon-1)(\epsilon+2)}{3V} \qquad \ldots\ldots(17)$$

and from (15) (Falkenhagen and Jacob 1951),

$$\left(\frac{\partial\epsilon}{\partial V}\right)_{\text{Onsager}}=-\frac{\epsilon(\epsilon-1)[2\epsilon+4n^2(\epsilon-1)-n^4]}{3[2\epsilon^2+n^4]V}. \qquad \ldots\ldots(18)$$

When these values for $\partial\epsilon/\partial V$ are substituted in (13), we see that

$$(\partial\gamma/\partial V)_{\text{Debye}}=0 \qquad \ldots\ldots(19)$$

and that

$$\left(\frac{\partial\gamma}{\partial V}\right)_{\text{Onsager}}=\frac{3(\epsilon^2-1)(\epsilon-n^2)}{2\pi(\epsilon+2)^2(2\epsilon^2+n^4)}. \qquad \ldots\ldots(20)$$

We thus conclude, on referring to equation (11), that no electrostriction occurs in a spherical specimen of dielectric according to Debye's theory. However, Onsager's theory does predict that electrostriction should occur in such a specimen when the possibility of orientational polarization exists (i.e. $\mu\neq0$).

§ 4. A MORE EXACT EVALUATION OF ΔP

The Helmholtz free energy F of a system is related to the partition function Z by the equation

$$F=-\beta^{-1}\log Z. \qquad \ldots\ldots(21)$$

In general, Z has the form

$$Z=\text{const.}\ T^{3/2}\int\exp[-\beta U]d\sigma, \qquad \ldots\ldots(22)$$

where U is the potential energy of the system for a particular configuration, and $d\sigma$ is an element of volume in configuration space.

From equation (9) we deduce that

$$\left(\frac{\partial F}{\partial V}\right)_{E^2,T}=-P-E^2\left(\frac{\partial\gamma}{\partial V}\right)_{E^2,T} \qquad \ldots\ldots(23)$$

and that

$$\left(\frac{\partial F}{\partial E^2}\right)_{V,T}=-\tfrac{1}{2}\gamma. \qquad \ldots\ldots(24)$$

Denoting the free energy of the system, in the absence of an external field, by F_0, we find

$$\left(\frac{\partial F}{\partial V}\right)_{E^2, T} - \left(\frac{\partial F_0}{\partial V}\right)_T = -P + P_0 - E^2\left(\frac{\partial \gamma}{\partial V}\right)_{E^2, T}. \qquad \ldots\ldots(25)$$

Since

$$P - P_0 = \Delta P, \qquad \ldots\ldots(26)$$

then by equation (11), equation (25) becomes

$$\left(\frac{\partial F}{\partial V}\right)_{E^2, T} - \left(\frac{\partial F_0}{\partial V}\right)_T = \Delta P. \qquad \ldots\ldots(27)$$

Our object now is to evaluate the left-hand side of equation (27), with the help of equation (22), and to investigate the nature of the approximations involved in equation (11) and also to indicate the reason why Debye's theory predicts that $\Delta P = 0$.

In order to calculate ΔP by means of equations (21), (22) and (27) it will be necessary to give an explicit form for U. This entails the adoption of a particular model. It is convenient to consider the cases of non-dipolar and dipolar dielectrics separately.

4.1. ΔP *for a Non-polar Fluid*

Consider a group of N molecules, of isotropic polarizability α. In the absence of an external field the partition function of the system, Z_0, is given by

$$Z_0 = \text{const.}\ T^{3/2} \int \exp\Big[-\beta \sum_{i>j}^{N} v(r_{ij})\Big] \prod_{k=1}^{N} d\tau_k, \qquad \ldots\ldots(28)$$

where $v(r_{ij})$ is the interaction energy of a pair of molecules, i and j, distance r_{ij} apart and $d\tau_k (= r_k^2 \sin\theta_k dr_k d\theta_k d\phi_k)$ is the element of volume occupied by molecule k. The presence of an external field will change Z_0 to Z such that

$$Z = \text{const.}\ T^{3/2} \int \exp\Big[-\beta\Big\{U_E + \sum_{i>j}^{N} v(r_{ij})\Big\}\Big] \prod_{k=1}^{N} d\tau_k. \qquad \ldots\ldots(29)$$

U_E is the increase in potential energy of the system for a particular configuration, resulting from the application of an external field.

If $\mathbf{f}_i$ denotes the total field acting on molecule i, then

$$U_E = \tfrac{1}{2}\alpha \sum_{i=1}^{N} f_i^2 - \alpha \sum_{i=1}^{N} \mathbf{f}_i \cdot \mathbf{E} - \tfrac{1}{2} \sum_{\substack{i,j=1 \\ (i\neq j)}}^{N} \mathbf{m}_i \cdot \mathbf{T}_{ij} \cdot \mathbf{m}_j, \qquad \ldots\ldots(30)$$

where $\mathbf{m}_i$ the induced dipole moment of the molecule i is given by

$$\mathbf{m}_i = \alpha \mathbf{f}_i. \qquad \ldots\ldots(31)$$

$\mathbf{T}_{ij}$ is a dyadic such that

$$\mathbf{T}_{ij} = -\nabla_i \nabla_j (1/r_{ij}) = r_{ij}^{-3}[(3\mathbf{r}_{ij}\mathbf{r}_{ij}/r_{ij}^2) - \mathbf{I}], \qquad \ldots\ldots(32)$$

where $\mathbf{I} (= \mathbf{ii} + \mathbf{jj} + \mathbf{kk})$ is the unit dyadic. Note that $\mathbf{T}_{ii} = 0$ (Kirkwood 1936). $\mathbf{T}_{ij} \cdot \mathbf{m}_j$ is the field acting on molecule i due to a dipole of magnitude m_j, induced in molecule j. Thus it is seen that

$$\mathbf{m}_i = \alpha \mathbf{E} + \alpha \sum_{j=1}^{N} \mathbf{T}_{ij} \cdot \mathbf{m}_j \qquad \ldots\ldots(33)$$

and therefore (Kirkwood 1936)

$$\mathbf{m}_i = \alpha \left\{ \mathbf{I} + \alpha \sum_{j}^{N} \mathbf{T}_{ij} + \alpha^2 \sum_{j,k}^{N} \mathbf{T}_{ij} \cdot \mathbf{T}_{jk} + \ldots \right\} \cdot \mathbf{E}. \qquad \ldots\ldots(34)$$

By means of equations (30), (31) and (34) we obtain

$$U_{\mathrm{E}} = -\tfrac{1}{2}\varphi E^2, \qquad \text{......(35)}$$

where

$$\varphi = \left\{ N\alpha + N\alpha^2 \sum_{j}^{N} \mathbf{e}\,.\,\mathbf{T}_{ij}\,.\,\mathbf{e} + N\alpha^3 \sum_{j,k}^{N} \mathbf{e}\,.\,\mathbf{T}_{ij}\,.\,\mathbf{T}_{jk}\,.\,\mathbf{e} + \dots \right\} \qquad \text{......(36)}$$

and where $\mathbf{e}$ is a unit vector parallel to $\mathbf{E}$, the external field. In view of (29) and (35), equation (21) takes the form

$$F = -\beta^{-1} \log \left\{ \text{const. } T^{3/2} \int \exp\left[-\beta \left(-\tfrac{1}{2}\varphi E^2 + \sum_{i>j}^{N} v(r_{ij}) \right) \right] \prod_{k=1}^{N} d\tau_k \right\}. \quad \text{..(37)}$$

If terms involving powers of E greater than the second are discarded, then

$$F = -\beta^{-1} \log \{ Z_0 [1 + \langle \tfrac{1}{2}\beta\varphi E^2 \rangle] \}. \qquad \text{......(38)}$$

The term $\langle \tfrac{1}{2}\beta\varphi E^2 \rangle$ is defined by the relation

$$\langle \tfrac{1}{2}\beta\varphi E^2 \rangle = \text{const. } T^{3/2} Z_0^{-1} \int \tfrac{1}{2}\beta\varphi E^2 \exp\left[-\beta \sum_{i>j}^{N} v(r_{ij}) \right] \prod_{k=1}^{N} d\tau_k. \qquad \text{......(39)}$$

Consequently

$$\left(\frac{\partial F}{\partial V} \right)_{E^2, T} = \left(\frac{\partial F_0}{\partial V} \right)_T - \tfrac{1}{2} E^2 \left(\frac{\partial \langle \varphi \rangle}{\partial V} \right)_T \qquad \text{......(40)}$$

and

$$\left(\frac{\partial F}{\partial E^2} \right)_{V, T} = -\tfrac{1}{2} \langle \varphi \rangle. \qquad \text{......(41)}$$

From equations (24) and (41) it will be seen that γ, the polarizability of a specimen of dielectric, is given by

$$\gamma = \langle \varphi \rangle = \left\{ N\alpha + N\alpha^2 \sum_{j}^{N} \langle \mathbf{e}\,.\,\mathbf{T}_{ij}\,.\,\mathbf{e} \rangle + N\alpha^3 \sum_{j,k}^{N} \langle \mathbf{e}\,.\,\mathbf{T}_{ij}\,.\,\mathbf{T}_{jk}\,.\,\mathbf{e} \rangle + \dots \right\}. \quad \text{..(42)}$$

This result is in agreement with the work of Yvon (1936, 1937) and Kirkwood (1936).

Substituting for $(\partial F/\partial V)_{E^2, T}$ in equation (27), from equation (40), we find

$$\Delta P = -\tfrac{1}{2} E^2 \frac{\partial \langle \varphi \rangle}{\partial V} = -\tfrac{1}{2} E^2 \left(\frac{\partial \gamma}{\partial V} \right)_{E^2, T}. \qquad \text{......(43)}$$

This shows that the expression obtained in §2, for ΔP, is correct as far as terms in E^2 are concerned.

In the case of a spherical specimen of a fluid dielectric it may be shown that (e.g. Böttcher 1952)

$$\langle \mathbf{e}\,.\,\mathbf{T}_{ij}\,.\,\mathbf{e} \rangle = 0, \qquad \text{......(44)}$$

and so

$$\Delta P = -\tfrac{1}{2} E^2 N\alpha^3 \sum_{j,k}^{N} \frac{\partial}{\partial V} (\langle \mathbf{e}\,.\,\mathbf{T}_{ij}\,.\,\mathbf{T}_{jk}\,.\,\mathbf{e} \rangle). \qquad \text{......(45)}$$

In the Lorenz–Lorentz theory it is assumed that γ is independent of the volume, and thus all except the first of the terms of equation (42) are ignored. Yvon and Kirkwood have shown that $\langle \mathbf{e}\,.\,\mathbf{T}_{ij}\,.\,\mathbf{T}_{jk}\,.\,\mathbf{e} \rangle$ differs from zero because the motions of the molecules are not statistically independent. In Kirkwood's phraseology one may say that *both* Debye and Onsager's theories neglect *translational* fluctuations. Kirkwood (1936) evaluated the expression

$$\sum_{j,k}^{N} \langle \mathbf{e}\,.\,\mathbf{T}_{ij}\,.\,\mathbf{T}_{jk}\,.\,\mathbf{e} \rangle$$

on the assumption that the molecules could be treated as hard impenetrable spheres of diameter a; the result he obtained was

$$\sum_{j,k}^{N} \langle \mathbf{e}\cdot\mathbf{T}_{ij}\cdot\mathbf{T}_{jk}\cdot\mathbf{e}\rangle = \frac{8\pi N}{3a^3V} - \frac{5\pi^2N^2}{3V^2}. \qquad \ldots\ldots(46)$$

For this particular model, therefore, we obtain, by combining equations (45) and (46),

$$\Delta P = \frac{E^2\pi N^2\alpha^3}{3V^2}\left(\frac{4}{a^3} - \frac{5\pi N}{V}\right). \qquad \ldots\ldots(47)$$

4.2. ΔP *for a Polar Fluid*

For simplicity the displacement polarization of the dielectric will be ignored, and it will be assumed that each molecule has a permanent dipole moment of magnitude μ. Z_0 is now given by

$$Z_0 = \text{const. } T^{3/2}\int \exp[-\beta\mathcal{U}_0]\prod_{k=1}^{N} d\tau_k d\Omega, \qquad \ldots\ldots(48)$$

where

$$\mathcal{U}_0 = -\sum_{i>j}^{N}[\boldsymbol{\mu}_i\cdot\mathbf{T}_{ij}\cdot\boldsymbol{\mu}_j + v(r_{ij})]. \qquad \ldots\ldots(49)$$

The term

$$-\sum_{i>j}^{N}[\boldsymbol{\mu}_i\cdot\mathbf{T}_{ij}\cdot\boldsymbol{\mu}_j]$$

is the dipole–dipole interaction energy of the system and $d\Omega$ is an element of $6(N-1)$ dimensional orientational configuration space.

In the presence of an external field E we have

$$Z = (2)^{-1}\text{const. } T^{3/2}\int \exp[-\beta(\mathcal{U}_E + \mathcal{U}_0)]\prod_{k=1}^{N} d\tau_k d\Omega \sin\Theta d\Theta \qquad \ldots\ldots(50)$$

where

$$\mathcal{U}_E = -ME\cos\Theta = -\sum_{i}^{N}\boldsymbol{\mu}_i\cdot\mathbf{E}. \qquad \ldots\ldots(51)$$

Adopting the same procedure as was used in § 4.1 and noting that F can only contain even powers of E, we obtain

$$F = -\beta^{-1}\log\{Z_0[1 + \tfrac{1}{6}\beta^2\langle M^2\rangle_P E^2]\}. \qquad \ldots\ldots(52)$$

$\langle M^2\rangle_P$ is defined by the relation

$$\langle M^2\rangle_P = \text{const. } Z_0^{-1}T^{3/2}\int M^2\exp[-\beta\mathcal{U}_0]\prod_{k=1}^{N} d\tau_k d\Omega. \qquad \ldots\ldots(53)$$

From equations (24), (27) and (52) it follows that

$$\Delta P = -\tfrac{1}{6}\beta E^2\left(\frac{\partial\langle M^2\rangle_P}{\partial V}\right)_{E^2,T} \qquad \ldots\ldots(54)$$

and that

$$\gamma = \tfrac{1}{3}\beta\langle M^2\rangle_P. \qquad \ldots\ldots(55)$$

We may write (Fröhlich 1949)

$$\langle M^2\rangle_P = N\boldsymbol{\mu}\boldsymbol{\mu}_{\text{sph}}^*, \qquad \ldots\ldots(56)$$

where $\boldsymbol{\mu}_{\text{sph}}^*$ is the dipole moment of a spherical specimen of dielectric when one of its constituent molecules is held in a fixed orientation.

According to Debye's theory the molecules of a dielectric are, effectively, assumed to be statistically independent, and therefore $\boldsymbol{\mu}\boldsymbol{\mu}_{\mathrm{sph}}{}^* = \mu^2$. This assumption leads to the result that $\Delta P = 0$, whereas Onsager's theory leads to the expression

$$\boldsymbol{\mu}_{\mathrm{sph}}{}^* = \left(\frac{3}{\epsilon+2}\right)\left(\frac{3\epsilon}{2\epsilon+1}\right)\boldsymbol{\mu}, \qquad \ldots\ldots(57)$$

which, when substituted in equation (54), leads to a value of ΔP which is different from zero.

It may be mentioned that the theory given by Frivold when applied to the case of a spherical specimen leads to a null result for ΔP, since he retained the Lorenz–Lorentz expression for the internal field (see § 5).

§ 5. DISCUSSION

The failure of the Lorenz–Lorentz and Debye theories to predict the occurrence of electrostriction in a spherical specimen of dielectric is fairly easy to understand. Both theories use the formula

$$\boldsymbol{\Phi} = [(\epsilon+2)/3]\boldsymbol{\mathfrak{E}}, \qquad \ldots\ldots(58)$$

to calculate the local or internal field, $\boldsymbol{\Phi}$. $\boldsymbol{\mathfrak{E}}$ is the potential gradient in the polarized dielectric. In the case of a sphere

$$\boldsymbol{\mathfrak{E}} = [3/(\epsilon+2)]\,\mathbf{E}, \qquad \ldots\ldots(59)$$

where $\mathbf{E}$ is the external field. Consequently in this particular case,

$$\boldsymbol{\Phi} = \mathbf{E}. \qquad \ldots\ldots(60)$$

This means that the field polarizing any one of the molecules is equal to the external field and is quite unaffected by the presence of the surrounding molecules. This result would only be correct if the dipole moments induced in the molecules were independent of the relative motions of the molecules. In other words the molecules are treated as though they were statistically independent. On the basis of this effective assumption of statistical independence, the result that $\Delta P = 0$ follows quite naturally.

Since we are only concerned with effects which are quadratic in E it is permissible to ignore any change in shape which might occur. For the same reason it was possible to neglect the effect of saturation when considering dipolar dielectrics.

We must now consider whether ΔP is positive or negative. It is possible to evaluate $(\partial\gamma/\partial V)_{E^2,T}$ from experimental data. For liquids $(\partial\gamma/\partial V)_{E^2,T}$ is usually found to be positive (Danforth 1931). However, it is found that for gases the sign of $(\partial\gamma/\partial V)_{E^2,T}$ depends on the density (see Böttcher 1952). For high densities $(\partial\gamma/\partial V)_{E^2,T}$ is positive and for low densities it is negative. This type of behaviour is predicted by the Yvon–Kirkwood theory. Equation (47) indicates that for sufficiently low densities an electric field gives rise to a positive pressure whereas at high densities the pressure is negative.

It is possible to extend the calculations in §4.1 so as to determine the values of the individual elements of the stress tensor P_{ij}; work to this end is in progress.

ACKNOWLEDGMENT

The author wishes to express his thanks to Mr. J. R. Pounder for many helpful discussions.

REFERENCES

ABRAHAM, M., and **BECKER, R.**, 1932, *The Classical Theory of Electricity and Magnetism* (London : Blackie).

BÖTTCHER, C. J. F., 1952, *Theory of Electric Polarisation* (Amsterdam : North-Holland Publishing Co.).

BRUHAT, M., and **PAUTHENIER, P.**, 1925, *J. Phys. Radium*, **6**, 1.

DANFORTH, W. E., 1931, *Phys. Rev.*, **38**, 1224.

DEBYE, P., 1929, *Polar Molecules* (New York : Chemical Catalogue Co.).

FALKENHAGEN, H., and **JACOB, H.**, 1951, *Ann. Phys., Lpz.*, **8**, 105.

FRIVOLD, O. E., 1921, *Phys. Z.*, **22**, 603.

FRÖHLICH, H., 1949, *Theory of Dielectrics* (Oxford : Clarendon Press).

GUGGENHEIM, E. A., 1949, *Thermodynamics* (Amsterdam : North-Holland Publishing Co.).

KIRKWOOD, J. G., 1936, *J. Chem. Phys.*, **4**, 592.

ONSAGER, L., 1936, *J. Amer. Chem. Soc.*, **58**, 1486.

YVON, J., 1936, *C.R. Acad. Sci., Paris*, **202**, 35; 1937, *Recherches sur la théorie Cinétique des Liquides: I, Fluctuations en Densité; II, La Propagation de la Lumière, Actualités Scientifiques et Industrielles, Nos.* 542 *and* 543 (Paris : Hermann et Cie).

Sublimation Nuclei

By S. C. MOSSOP
Clarendon Laboratory, Oxford†

Communicated by G. M. B. Dobson; *MS. received 25th July* 1955

Abstract. The formation of ice crystals in cloud chambers generally takes place by the freezing of supercooled water drops rather than by the sublimation of vapour on to suitable nuclei. Fournier d'Albe found that particles of cadmium iodide possessed the unusual property of allowing ice crystals to grow on their surfaces by sublimation, under certain circumstances. Three other substances, $CaCO_3$, $CaSO_4 . 2H_2O$ and sodium bentonite clay have now been found to behave similarly. An explanation is suggested.

§ 1. Introduction

Cloud chamber experiments on outdoor air show that ice crystals form only when expansions large enough to produce a water cloud are made (Fournier d'Albe 1949). Such crystals are thus formed by the freezing of water drops rather than by the sublimation of vapour on to suitable 'sublimation nuclei'. This was confirmed when aerosols composed of a wide variety of chemicals were introduced into the expansion chamber in place of outdoor air.

However, Fournier found that one salt, cadmium iodide, behaved differently from the rest. A dilute solution was sprayed into the cloud chamber and expansions were made to give cloud at a series of decreasing temperatures. The first ice formation occurred at approximately −41°c which is now considered to be the threshold temperature for spontaneous nucleation of water drops of the size formed in Fournier's chamber (Mason 1952). When the chamber was allowed to warm up, it was found that very small expansions caused ice crystals to grow on some of the remaining particles; instead of appearing only at the dew point, crystals were apparently formed by sublimation. This effect continued as the cloud chamber was warmed up until a temperature of −9·4°c was reached; above this it was no longer observed.

Fournier's experiments on cadmium iodide have now been repeated and extended to other substances which display similar properties.

§ 2. Present Work

2.1. *Cadmium Iodide*

Several experiments were made on cadmium iodide with Fournier's cloud chamber (1949) and his findings were, in general, confirmed. Aerosols were formed both by spraying solutions (concentrations N/5 to N/250), and by heating a small amount of the salt on a platinum wire to give a visible smoke.

When a cloud was made by condensation upon these particles at successively lower temperatures the threshold for the first appearance of ice crystals, in a concentration of one crystal per thousand droplets, was approximately $-24°$c. Below this there was sparse ice crystal formation down to approximately $-41°$c, where the usual spectacular increase in the number of crystals took place. That Fournier did not observe any ice formation between -24 and $-41°$c indicates that he must have worked with a much lower concentration of aerosol particles.

---

† Now at South African National Physical Laboratory, Pretoria.

Once ice crystals had been formed below $-24°$c it was found that after the cloud had evaporated subsequent small expansions caused some of the brighter particles to grow into ice crystals. These expansions were much smaller than that required to produce water droplets. The crystals could be kept in the field of view and the ice could be caused to grow or diminish at will by small expansions or compressions. The number of such 'sublimation nuclei' was greatly increased after expansions had been made past the $-41°$c threshold; as many as 50 per 1000 cadmium iodide nuclei were then observed.

The present observations confirm Fournier's discovery that as the chamber is warmed up the 'sublimation nuclei' persist only until about $-10°$c, and thereafter they cease to act in this way. After such warming no ice crystals can then be seen, even with large expansions, until the cloud temperature is again reduced below $-24°$c.

The temperature above which this sublimation effect became inoperative varied from $-9{\cdot}1°$c for N/5 solutions, to $-9{\cdot}7$ for N/250 solutions and $-12{\cdot}3$ for 'smokes'. The fact that smoke particles are, in general, smaller than those produced by spraying a solution, suggests that there may be some correlation between this temperature and particle size.

It was found that a small but definite expansion was necessary for this type of crystal growth. At $-14°$c this minimum expansion produced a saturation of 107% with respect to ice or 92% with respect to water. At $-27°$c the percentage saturations were 110 and 84 respectively.

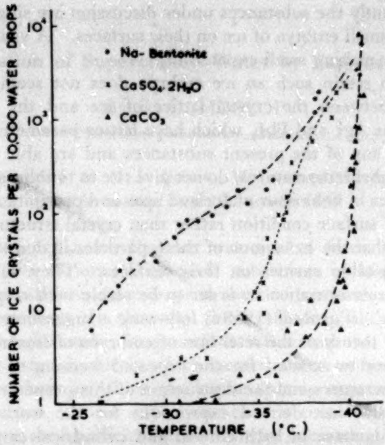
If the chamber was warmed from $-14°$ directly to $-6°$c without making intermediate expansions, no 'sublimation nuclei' were then observed. The 'sublimation nuclei' could not have been removed by sedimentation, but rather became inactive due to warming above a critical temperature.

2.2. *Other Instances of the Formation of 'Sublimation Nuclei'*

In the course of investigations of the nucleating properties of other substances, a careful check was made for the appearance of similar 'sublimation nuclei'. The phenomenon appears to be quite rare and only the following cases were discovered out of more than 50 materials tested.

(*a*) $CaCO_3$. A sample of Iceland spar of 'analytical purity' (certainly less than 0·1% impurity) was tested by spraying a water-suspension into the cloud chamber and then cooling. The effectiveness of the particles as freezing nuclei is illustrated in the figure. On warming up from $-41°$c, between 1% and 5% of the particles were found to act as 'sublimation nuclei'. With a total of 500 particles in the field of view, the 'sublimation nuclei' were last observed at a temperature of $-3{\cdot}7°$c, when the expansion necessary for ice crystal formation gave rise to a faint cloud. In subsequent experiments with lower particle concentrations they disappeared earlier.

It was noteworthy that an 'Analar' sample of $CaCO_3$, presumably made by a precipitation process, did not give rise to 'sublimation nuclei'.



(*b*) *CaSO$_4$.2H$_2$O (Gypsum)*. 'Analar' material was used and its ice forming properties are indicated in the figure. The concentration and behaviour of 'sublimation nuclei' were generally as described for Iceland spar. The highest temperature at which they were observed was −5·3°c, the last expansion again being enough to produce faint cloud formation. In one experiment, after sublimation crystals had been observed at −9°c, the chamber was warmed to +1°c without intermediate expansions, and then cooled again to −11°. No crystals could then be seen on making expansions.

(*c*) *Sodium Bentonite Clay*. A sprayed suspension of these particles gave freezing nuclei as shown in the figure. About 5% of the particles acted as sublimation nuclei from −41 to −6·5°c.

§ 3. Discussion

Fournier (1949) suggested that the cadmium iodide sublimation nuclei may have been small frozen particles of saturation solution, which melted on passing the eutectic point, though why they did not freeze during the initial cooling was not clear.

Mason (1950) offered the more convincing explanation that warming up to temperatures below −9°c caused only partial evaporation of the ice crystals and that a microscopic oriented film of ice was retained on the surface of the CdI_2 particles.

This is supported by the results of the present experiments, particularly the facts, firstly that 'sublimation nuclei' are observed when, and only when, ice crystals have been formed by the usual freezing mechanism, and secondly that 'sublimation nuclei' lose their peculiar property when warmed above 0°c and re-cooled.

Usually when ice crystals are formed in a cloud chamber expansion, they will either grow large enough to fall to the bottom and be lost, or else the ice will evaporate in the subsequent recompression, leaving only the original freezing nucleus. Apparently the substances under discussion are singular in that some particles retain a small embryo of ice on their surfaces. A visible ice crystal can then be grown by making small expansions.

The ability to retain such an ice embryo does not seem to depend upon closeness of fit between the crystal lattice of ice and the substrate. Thus substances such as AgI and PbI_2 which have lattice parameters much closer to those of ice than any of the present substances and are able to act as freezing nuclei at much higher temperatures, do not give rise to 'sublimation nuclei' at all. Also the differences in behaviour of Iceland spar and precipitated calcite point to the importance of surface condition rather than crystal structure.

It is possible that the behaviour of these particles is due to the retention of ice embryos in suitable cavities on their surfaces. They will then no longer require a slight supersaturation in order to be stable with respect to the ice on the chamber walls. Turnbull (1950 a) following a suggestion of Volmer (1939) has used a similar theory of the retention of embryos of frozen metal in cavities in a foreign surface to account for the repeated freezing of molten metals at reproducible temperatures and the destruction of this property by superheating.

Turnbull (1950 b) has derived expressions for the work of formation of embryos of a solid phase in both conical and cylindrical cavities in a foreign substrate. He shows that an embryo will be retained up to a critical temperature which depends on the dimensions of the cavity and the interfacial surface energies involved. Unfortunately we have no knowledge of the interfacial energies involved in the present instances, otherwise it would be possible to calculate the size of the cavities assumed of simple shape in our 'sublimation nuclei'. However, Turnbull's theory does give a plausible explanation for the disappearance of their ice-forming ability after the temperature has risen past a certain point.

In the case of cadmium iodide, there is another possible explanation for the melting of the ice embryo, since it will be unstable above the eutectic point of CdI_2 ($-8{\cdot}5°$c for bulk materials) and a droplet of solution will result, as Mason has pointed out.

The meteorological importance of this phenomenon is probably slight. No observations on outside air either by Fournier or the present author have disclosed any 'sublimation nuclei' and they are likely to be of rare occurrence. Their singular property might occasionally enable ice particles from cirrus clouds to persist under conditions where they would normally disappear by sublimation, and then to grow again on entering air saturated with respect to ice.

Acknowledgments

I wish to express my thanks to Professor Lord Cherwell for the use of the facilities of the Clarendon Laboratory, and I am deeply indebted to Professor G. M. B. Dobson under whose supervision the present work was performed, for continued advice and encouragement. I wish to thank the H. B. Webb Trustees, South Africa, for the grant of a Research Scholarship.

References

Fournier d'Albe, E. M., 1949, *Quart. J. R. Met. Soc.*, **75,** 1.
Mason, B. J., 1950, *Quart. J. R. Met. Soc.*, **76,** 59; 1952, *Ibid.*, **78,** 22.
Turnbull, D., 1950 a, *J. Appl. Phys.*, **21,** 1022; 1950 b, *J. Chem. Phys.*, **18,** 198.
Volmer, M., 1939, *Kinetik der Phasenbildung* (Leipzig : Steinkopff).

The Nucleation of Supercooled Water by Various Chemicals

By S. C. MOSSOP
Clarendon Laboratory, Oxford†

*Communicated by G. M. B. Dobson; MS. received 25th July 1955
and in amended form 21st October 1955*

Abstract. A cooled expansion chamber was used to form water clouds upon aerosol particles of a variety of chemical substances of low solubility. The proportion of ice crystals to water drops in the clouds was observed at a series of decreasing temperatures. In this way the ice-forming ability of the various chemicals could be assessed. In general it was found that for the particular substances chosen, the chemicals which caused water to freeze at the highest temperatures were those in which the lattice spacing approached most closely that in the ice lattice. Some exceptions to this rule show that geometrical considerations alone are insufficient to account for nucleating ability.

§ 1. Introduction

It has long been realized that the freezing of supercooled liquids can be initiated by impurities. The reasons for this action are not fully known. The purpose of the present work is to investigate the nucleation of supercooled water by a number of chemical substances, and to attempt to relate their 'effectiveness' to other physical and chemical properties. Here one substance is designated a more 'effective' or 'potent' nucleating agent than another if it causes supercooled water drops to freeze at a higher temperature.

§ 2. Previous Work

When a cloud of water drops formed in an expansion chamber is cooled below approximately −40°c, there is a vast increase in the number of ice crystals present (Schaefer 1946, Cwilong 1947). Because this threshold temperature is the same for drops formed upon many different types of condensation nuclei, this transition is regarded as a property of the water itself, i.e. it is considered that spontaneous nucleation takes place (Mason 1952). If droplets of similar size containing a foreign substance are found to freeze at a higher temperature when cooled at the same rate, then one is justified in assuming that the impurity has assisted freezing by the provision of a 'freezing nucleus'.

Many materials have been found which assist the nucleation of supercooled water to a greater or lesser degree. Investigation of the effect of various chemicals on the freezing points of water drops have been made by Vonnegut (1947), Fournier d'Albe (1949), Lafargue (1950) Reynolds (1950), Aufm Kampe and Weickmann (1951), Hosler (1951), Bigg (reported by Mason 1953), and Hosler (1954).

Vonnegut showed that heated silver iodide produced a 'smoke' of very tiny particles, which caused copious ice crystals when injected into a water cloud below a temperature of −4°c. He ascribed this action to the resemblance between the crystal structures of silver iodide and ice. However, lead iodide, which also has a crystal structure very similar to that of ice, was found to cause the freezing of water drops only at −20°c.

---

† Now at South African National Physical Laboratory, Pretoria.

The influence of crystal structure was discounted by Hosler (1951) who found that, of approximately 50 chemicals tested, the most effective for ice formation were AgI, $AgNO_3$, Ag_2S and PbI_2 (acting at temperatures in the region of −4°c), while many iodides were found to act as freezing nuclei from −12°c downwards. Not all these materials belong to the hexagonal system, nor do they all have lattice dimensions similar to those in ice. Hosler therefore attempted to explain his results in terms of a hypothesis of Weyl (1951) that the addition of highly polarizable ions or molecules to a water drop will reduce the amount of supercooling necessary to form ice crystals. Unfortunately, owing to lack of information on the polarizability of the compounds used, this theory could not be adequately tested.

In Hosler's later paper (1954) he gives rather lower temperatures than those mentioned above for the initial appearance of ice crystals for the chemicals mentioned. He also reports on the effect of various solutes on the freezing point of water. Some salts lowered the freezing point in all concentrations; others raised the freezing point, up to a certain concentration, above which a decline again set in.

Bigg (reported by Mason 1953) experimented on droplets of various solutions suspended at the interface between two immiscible liquids. Working with a concentration of one solute molecule to 1000 water molecules (i.e. [illegible] molar) he also showed that a wide variety of dissolved salts could cause an elevation of the freezing point of water drops over the value found for 'pure' water. The suggested explanation follows Mason's idea that freezing may be initiated by the aggregation of water molecules round solute ions (Mason 1950).

We see that although there has been a considerable accumulation of experimental data on the threshold temperatures at which ice formation is first initiated by various chemicals, no full explanation of the action of freezing nuclei is yet in sight. There are considerable discrepancies between the results of different workers on the same material. These inconsistencies may be due to differences in the concentration of the aerosol particles used, to differences in the size and surface properties of particles, caused by the method of dispersion, or to differences in the sizes of the water drops and in the sensitive time of the cloud chambers used. There is as yet little information on the effect of these variables.

§ 3. PRESENT EXPERIMENTAL METHOD

The present series of experiments was intended to take advantage of the potentialities of the cloud chamber designed by Fournier d'Albe (1949). This had an observation system which enabled individual cloud drops and ice crystals to be observed and counted. Thus quantitative measurement could be made of the number of freezing nuclei active at a particular temperature in relation to the total number of nuclei present in the chamber. A valid comparison could then be made between the effectiveness of nuclei of different chemical compositions.

The choice of the substances whose effectiveness was to be compared was governed by the following considerations:

(i) As we have seen above, freezing of water can be assisted either by the presence of a foreign solid substrate on which the new ice crystal can grow, or else by the presence of suitable solutes in the water. (It is conceivable that there are moderately soluble substances with which both mechanisms come into play.) Different experimental techniques are required to investigate the two processes. For the former we wish to study the freezing of water drops formed upon individual nuclei, while for the latter we need to be able to control and measure the concentration of solute in water drop. Cloud chamber techniques are thus of limited application in the latter case. The present investigation was therefore confined to low-solubility materials in an attempt to ensure that nucleation was caused by the nuclei in solid form.

(ii) For ease in interpreting results, the experiments were confined to simple inorganic salts of known crystal structure.

(iii) In an attempt to obtain reproducibility of results, almost all the materials were introduced into the cloud chamber in the form of 'smokes' produced by heating the salt upon a wire to the lowest temperature that would give a visible smoke. The choice of salts was restricted to those which do not decompose when heated in this way.

§ 4. EXPERIMENTAL PROCEDURE

4.1. *The Cloud Chamber*

The chamber was designed and used by Fournier d'Albe (1949) and a brief description will suffice.

It consisted of a brass cylinder of inside diameter 3 in. and height 12 in. A system of lenses was used to focus a powerful beam of light passing down the tube at a point near the centre, so that a volume of about 0·1 cm^3 was intensely illuminated. Any suspended particle in this volume could be examined through a microscope at right angles to the beam. Above and below the observation space were diaphragms 1 inch apart which reduced the stray light.

The chamber could be cooled down to -40°c, and was connected to a bellows system by which expansions of known ratio could be made.

4.2. *Determination of Temperature produced by Expansion*

The temperature of the brass wall next to the observation space could be measured by a thermocouple system. From this it was possible to calculate the air temperature produced by an expansion of known ratio, allowing for the non-adiabatic nature of the expansion according to the method used by Fournier (1949). In the largest expansions used it is estimated that the final air temperature could be measured accurately to $\pm 0{\cdot}4$°c.

4.3. *Production of Clean Air*

Before an experiment the chamber was first cleaned by flushing it with air from the compressed air line which had been bubbled through distilled water and then passed through a cotton wool filter. This technique, described by Wylie (1953), was found to get rid of all but a very few suspended particles (less than 1 per cm^3).

4.4. *Production of Aerosols*

In most cases the test substance was vaporized on an electrically heated platinum wire, and introduced into the chamber in a stream of clean air.

It was first necessary to verify that the wire itself did not give rise to any freezing nuclei. When heated to a red heat, a platinum coil gave vast numbers of condensation nuclei. Water drops formed upon these nuclei froze only at the '−40°c threshold' and thus did not contain any freezing nuclei. This confirms the findings of Fournier d'Albe (1949).

However, when a platinum wire was heated above 1000°c and clouds were formed upon the resultant nuclei at successively lower temperatures, ice crystals appeared sparsely at all temperatures from −22°c downwards. The same results were obtained when the experiment was repeated in nitrogen or when the wire was initially cleaned in boiling nitric acid. These freezing nuclei were not found however, if the wire had previously been heated to 1450°c for an hour, and thus apparently consisted of some volatile material.

Similar properties were shown by the nuclei obtained when wires of nichrome, tungsten, nickel and platinum–10% iridium were heated to 1000°c.

In the following experiments, small quantities of chemicals were heated upon coils of platinum wire at the lowest temperature that would give a visible 'smoke'. Whenever a temperature approaching 1000°c was necessary, the wire was first heat-treated as described above to get rid of all freezing nuclei.

A different technique was used in the production of MgO aerosol. A coil of magnesium ribbon was ignited in a crucible, and the smoke was carried into the cloud chamber with a stream of clean air. To prevent the intense heat of the burning magnesium producing nuclei from the crucible material, the ribbon was partially severed so that it stopped burning before the flame could touch the porcelain.

4.5. *Estimation of Particle Size*

From the brightness of individual particles it was possible to divide them into four rough size categories: large, medium, small and 'invisible'. The last mentioned were only rendered visible when water drops grew upon them. The approximate diameters of particles corresponding to these categories were estimated from the work of Fournier d'Albe (1949) and from some electron micrographs of CdI_2 smoke kindly made by Dr. Lees of the Atomic Energy Research Establishment, Harwell.

Table 1. Aerosol Particle Sizes

	Diameter
Large	$>0{\cdot}5\ \mu$
Medium	$0{\cdot}5$ to $0{\cdot}2\ \mu$
Small	$0{\cdot}2$ to $0{\cdot}05\ \mu$
'Invisible'	$<0{\cdot}05\ \mu$

4.6. *Expansion Chamber Experiments*

The aerosol was passed into the cloud chamber and diluted with clean air to the desired particle concentration (up to 5000 particles in the field of view). The concentration could be estimated using diaphragms of various sizes at the image plane between the microscope objective and eyepiece to reduce the number of visible particles to a known fraction of the total. As has been mentioned above, not all the particles were large enough to be initially visible, and it was therefore necessary to make an expansion to cloud point so that they could grow and be counted. Approximately 20 such counts would be made for each determination of concentration. Since the cloud drops did not remain stationary but moved downwards shortly after the expansion owing to the formation of convection cells, this involved a rapid estimate of the number of pinpoints of light visible in a small circular opening. This might vary between 0 and 10. With practice and by taking the average of many observations it was thought that the concentration of nuclei could be accurately gauged to within a factor of 2.

The chamber was cooled, and at any desired temperature down to −38°c a series of expansions of increasing ratio was made. All the particulate materials used in the present investigation were found to allow cloud formation by rapid condensation of water when relative humidities near to or greater than 100% with respect to water were reached in an expansion. Below a certain air temperature, depending on the nature of the aerosol particles being used as condensation nuclei, ice crystals would also be seen. As Fournier d'Albe found, these occurred only in the company of water drops and not at lower expansions, i.e. they were produced by freezing of water drops and not by sublimation. From the chamber temperature and the expansion ratio, the air temperature for the first production of ice crystals could be found. Further expansions were made at lower temperatures and the number of ice crystals noted. These rapid visual estimates of the number of ice crystals in the field of view were probably accurate to a factor of 2.

There was no difficulty in distinguishing ice crystals from cloud droplets or aerosol particles. Ice crystals were much brighter than droplets, fell faster and twinkled as they fell.

Each time a cloud was made, the expansion was maintained for several minutes afterwards so that any ice crystals would grow and fall to the bottom of the chamber. This minimized the risk of their being counted again at the next lower temperature.

An experimental run was continued until either the '−40° threshold' was reached, when the usual vast increase in the number of ice crystals took place, or until crystals were formed too copiously for accurate counting, at some temperature above −40°. The chamber was then allowed to warm up and small expansions were made to see if the test particles acted as 'sublimation' nuclei in the manner described for CdI_2 by Fournier d'Albe. None of the materials in the present investigation displayed this property.

For each substance tested, at least three experimental runs were made with different particle concentrations.

§ 5. EXPERIMENTAL RESULTS

In table 2 are listed the substances whose ice-forming properties have been investigated in this way. The choice of materials for these experiments was governed by the considerations mentioned in § 3, which severely limit the number of substances that can be used.

The solubilities given in table 2 are taken from Seidell (1940) and the crystallographic information from Wyckoff (1951).

Table 2. Solubility, Nucleation Threshold and Crystal Structure of Various Artificial Freezing Nuclei

Substance	Solubility (gm/100 cm^3)	Temperature (°c) for 1 ice crystal per 10 000 drops	Crystal system	Crystal structure	Lattice parameters a_0	Lattice parameters c_0	Δa_0	$\delta = \frac{\Delta a_0}{a_{0(\mathrm{ice})}}$	Particle sizes
Ice			Hex.		4·51	7·35			
AgI	8×10^{-6}	−6	Hex.	ZnO	4·58	7·49	0·07	0·015	S, I
PbI_2	4×10^{-2}	−7	Hex.	$Cd(OH)_2$	4·54	6·86	0·03	0·007	M, S, I
HgI_2	3×10^{-3}	−17	Tetrag.		4·36	12·34	0·15	0·033	L, M, S, I
Hg_2Br_2	4×10^{-6}	−26	Tetrag.		4·65	11·10	0·14	0·031	
TlI	6×10^{-3}	−28	Cubic.		4·20		0·31	0·07	
MgO	6×10^{-4}	−30	Cubic.	NaCl	4·20		0·31	0·07	L, M, S, I
CdS	$1{\cdot}3 \times 10^{-4}$	−36	Hex.	ZnO	4·13	6·69	0·38	0·085	
HgS	1×10^{-6}	−40	Hex.		4·16	9·54	0·35	0·078	
ZnS	$1{\cdot}5 \times 10^{-7}$	−40	Hex.	ZnO	3·81	6·23	0·70	0·15	M, S
NiS	4×10^{-4}	−40	Hex.	NiAs	3·42	5·30	1·09	0·24	M, S

The results of the cloud chamber experiments are displayed graphically in figures 1, 2 and 3, as 'freezing nucleus spectra' for the various salts used, i.e. cumulative totals of the number of ice crystals observed at successively decreasing temperatures, as a fraction of the total number of water drops. Thus at any temperature one may read off the proportion of nuclei of a particular aerosol which act as freezing nuclei at or above that temperature. As an aid in doing this, curves have been fitted by eye to the experimental points. The temperatures for which the ice crystal concentration reaches 1 per 10,000 drops are given in table 2.

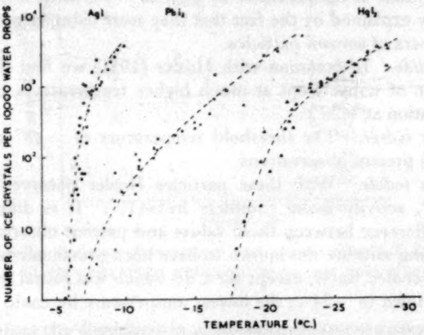


Figure 1. Relation between temperature and the proportion of ice crystals in water clouds formed by condensation upon aerosol particles of AgI, PbI_2, HgI_2.

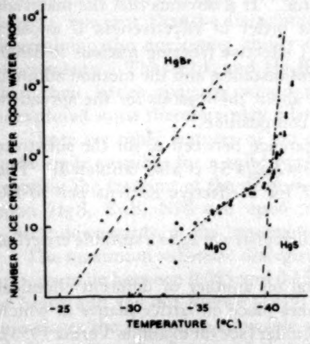


Figure 2. Relation between temperature and the proportion of ice crystals in water clouds formed by condensation upon aerosol particles of Hg_2Br_2, MgO, HgS.

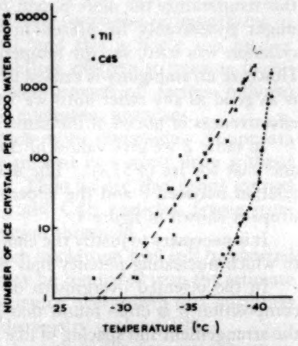


Figure 3. Relation between temperature and the proportion of ice crystals in water clouds formed by condensation upon aerosol particles of TlI, CdS.

The freezing nucleus spectra for ZnS and NiS have not been given since they are virtually identical with that for HgS, namely the '−40°c spontaneous nucleation spectrum', showing that these particles play no part in the freezing process.

§ 6. Discussion of Results

6.1. *Comparison with Results of Previous Workers*

(*a*) *Silver iodide.* Present results show fair agreement with those of Vonnegut (1947), Aufm Kampe and Weickmann (1951) and Hosler (1951). These workers observed ice crystals at temperatures as high as −3°c with these nuclei, which may be partially explained by the fact that they were using large cloud chambers with vast numbers of aerosol particles.

(*b*) *Lead iodide.* In common with Hosler (1951) we find that this salt can cause nucleation of water drops at much higher temperatures than Vonnegut's original observation at −20°c.

(*c*) *Mercuric iodide.* The threshold temperature of −18° given by Hosler agrees well with present observations.

(*d*) *Thallous iodide.* With these particles Hosler observed ice crystals to form at −13°c, activity being complete by −17°. It is difficult to account for the large difference between these values and present observations.

The remaining salts are not known to have been previously tested as freezing nuclei for supercooled water, except for CdS which was found by Hosler to give no ice crystals down to −24°c, the lowest temperature he could reach.

6.2. *Relation between Nucleating Potency and Crystal Structure*

As a measure of the 'effectiveness' or 'potency' of the various nucleating materials, we have chosen the temperature at which the proportion of effective freezing nuclei to the total number of nuclei reaches 1 per 10 000. The higher this temperature the more potent the nucleator. It is obvious that the materials might conceivably be placed in a different order of effectiveness if another criterion was used, say the temperature for 1 effective freezing nucleus in 1000. However no ambiguity is caused in the present instance, and the method adopted is as good as any other until we know more about the reasons for the spread in effectiveness of nuclei of the same chemical composition.

In table 2 we give values of Δa_0, the difference between a_0 for the substrate and that for ice (4·51 Å). The disregistry $\delta = \Delta a_0/4{\cdot}51$ is also tabulated. The relation between δ and the supercooling ΔT for 1 effective nucleus per 10 000 drops is shown in figure 4.

It is necessary to justify the choice of the disregistry in a_0 as a suitable criterion to which nucleating potency may be related.

In the oriented overgrowth of one crystal on another of different chemical composition it is often found that growth takes place on lattice planes in which the arrangement and spacing of like atoms is similar (see for example Verma 1953). In the salts of hexagonal structure that we have used the basal planes are similar to that in ice, the (−) ions occupying the position of alternate oxygens in ice. The spacing between these corresponding atoms is given by a_0 in each case, so that Δa_0 is the misfit.

In seeking an explanation of the ability of the non-hexagonal salts HgI_2, Hg_2Br_2, TlI and MgO to nucleate supercooled water, it is necessary to examine their structure to find possible lattice planes where the atomic arrangement is similar to that in a plane in the ice structure. A search has failed to reveal any such planes. It is therefore necessary to postulate that nucleation is due to similarities in spacings between atoms in single rows and not to two-dimensional similarity. It so happens that the closest resemblance in such rows is that given by the a_0 spacing in the basal plane in each case, as compared with the spacing of 4·51 Å between oxygen atoms in the basal plane of ice.

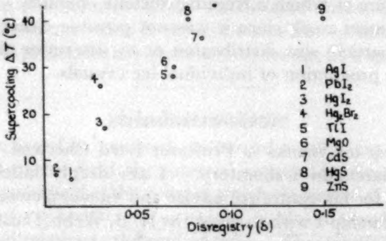


Figure 4. Relationship between disregistry δ and supercooling ΔT for production of 1 ice crystal per 10 000 water drops condensed on various types of aerosol particle.

For these reasons the disregistry in a_0 has been chosen as a parameter to which nucleating ability may be related. There will undoubtedly be cases where the disregistry in some other dimension would be more suitable.

From figure 4 there appears to be a definite relationship between ΔT and δ, the best nucleating materials having, in general, the smallest disregistry. This is in agreement with ideas put forward by Turnbull and Vonnegut (1952). They suggest, however, that the disregistry is not the only parameter of importance in determining the nucleating ability of a crystal, and this is borne out by our experiments. Thus PbI_2 and Hg_2Br_2 are much less effective as freezing nuclei than their lattice spacings would suggest. Non-geometrical factors hitherto unexplored must therefore play a large part in nucleation processes.

There are other features of our results which merit discussion. It appears to be quite possible for a solid nucleus to be present in a water drop without assisting the freezing of the drop in any way. Thus we see that drops formed upon HgS, ZnS, NiS and upon most MgO and CdS particles, freeze from −40°c downwards, at the 'spontaneous nucleation threshold'.

The maximum substrate disregistry for which nucleation may still be assisted appears to lie between 0·085 and 0·15. This may be compared with the estimate of Frank and van der Merwe (1949) that for oriented overgrowth of a misfitting monolayer the maximum allowable substrate disregistry is about 10%.

From these results we see that it may not be impossible for water drops in solid containers to freeze without the walls playing any part in the process.

Besides the differences in nucleating ability from one salt to another, there is a vast range in the effectiveness of particles of the same material. There may be differences of more than 10°c in the temperatures at which two particles of PbI_2, for instance, act as freezing nuclei. This wide spread of 'freezing nucleus spectra' has gone without comment in the literature.

There is obviously a wide variation in the effectiveness of nucleation sites from one particle to another. This may be purely a size effect, since nucleation can only take place if the site is large enough to accommodate an embryo of the new phase larger than the critical size appropriate to that temperature (Turnbull 1950). It is to be expected that the area of sites parallel to the most favourable lattice plane for nucleation would vary with particle size and with the conditions under which the particle grew.

This can only be tested by further experiments on the effect of particle size upon the temperature at which a freezing nucleus operates. Little guidance is afforded by the present work since it was not possible either to make accurate measurements of particle size distribution or to determine the size of nucleus responsible for the production of individual ice crystals.

ACKNOWLEDGMENTS

I wish to express my thanks to Professor Lord Cherwell, for the use of the facilities of the Clarendon Laboratory. I am deeply indebted to Professor G. M. B. Dobson, for his continued advice and encouragement. For the grant of a Research Scholarship I wish to thank the H. B. Webb Trustees, South Africa.

REFERENCES

AUFM KAMPE, H., and WEICKMANN, H. K., 1951, *J. Met.*, **8,** 283.
CWILONG, B. M., 1947, *Proc. Roy. Soc.* A, **190,** 137.
FOURNIER D'ALBE, E. M., 1949, *Q. J. Roy. Met. Soc.*, **75,** 1.
FRANK, F. C., and VAN DER MERWE, J. H., 1949, *Proc. Roy. Soc.* A, **198,** 216.
HOSLER, C. L., 1951, *J. Met.*, **8,** 326; 1954, *Proceedings of the Toronto Meteorological Conference*, 1953 (London: Royal Meteorological Society), p. 253.
LAFARGUE, C., 1950, *Centenary Proc. Roy. Met. Soc.*, 61.
MASON, B. J., 1950, *Quart. J. Roy. Met. Soc.*, **76,** 59; 1952, *Ibid*, **78,** 22; 1953, *Archiv. Met. Wien*, **6,** 1.
REYNOLDS, S. E., 1950, *Signal Corps Res. Proj.* 399–07–022, *Progress Rep. No.* 6, New Mexico School of Mines, Socorro, U.S.A.
SCHAEFER, V. J., 1946, *Science*, **104,** 457.
SEIDELL, A., 1940, *Solubilities* (New York: Van Nostrand.)
TURNBULL, D., 1950, *J. Chem. Phys.*, **18,** 198.
TURNBULL, D., and VONNEGUT, B., 1952, *Ind. and Eng. Chem.*, **44,** 1292.
VERMA, A. R., 1953, *Crystal Growth and Dislocations* (London: Butterworths).
VONNEGUT, B., 1947, *J. Appl. Phys.*, **18,** 593.
WEYL, W. A., 1951, *J. Colloid. Sci.*, **6,** 389.
WYCKOFF, R. W. G., 1951, *Crystal Structures* (New York: Interscience).
WYLIE, R. G., 1953, *Proc. Phys. Soc.* B, **66,** 241.

The Electrostatic Centre of a Conductor[†]

By R. CADE and D. O. VICKERS[‡]
University College of the West Indies, Jamaica

Communicated by R. Fürth; *MS. received 2nd June* 1955
and in amended form 27th July 1955

Abstract. It is shown that relative to a conductor influenced by a uniform electric field, there exists a fixed point with special properties. If the conductor has no total charge, its potential is always the same as that of the uniform field alone at this point. When the moment of the electrostatic forces on the conductor is taken about the point, the value is the same whether or not the total charge is zero. The general principles are illustrated by results on a special problem.

§ 1. INTRODUCTION

It is well known that if an insulated, uncharged, conducting sphere is introduced into a uniform electric field, the potential acquired by the sphere is the same as that of the original field at the point at which the centre of the sphere is placed. The same is true of a spheroid, and moreover, holds independently of the orientation of the spheroid relative to the applied field. The question may thus be asked, is there, relative to *any* conductor, a fixed point, independent of the orientation of the conductor, such that when the conductor is influenced by a uniform field and has no total charge, its potential is the same as that of the applied field alone at that point? We shall show that there is. Further, we shall show that if the conductor has a net charge, the total moment of the electrostatic forces upon it about any axis through this point is the same as if the total charge were zero. In other words, in the language of elementary statics, when the action on the conductor is represented as a single force together with the couple that is present when there is no total charge, the force always acts through the point. The point is clearly analogous, in some respects, to the centre of gravity, but its relationship to a charge distribution is essentially less simple than that of the centre of gravity to a mass distribution on account of the mobility of charge, a reorientation of a conductor causing a redistribution of charge so as to maintain constant potential. We propose, for the want of something more suitable, the name *electrostatic centre* for the point.

The electrostatic centre is of an elementary character as an electrostatic property, and it has obvious practical significance to such matters as the design of electrical instruments. Nevertheless, a search of the literature available to us has failed to reveal that it has been discovered before. A possible reason is that, like the centre of gravity, it is distinct from the geometrical centre only in the case of an asymmetrical body and, on account of mathematical difficulty, very little work has been done on the solution of potential problems involving asymmetrical conductors. It was, in fact, during the investigation of such a problem by one of us (D.O.V.) that the electrostatic centre came to light, more or less by accident. After the general theory, a résumé of this calculation will be given by way of an example.

[†] § 3 of this paper is part of a thesis by D. O. Vickers accepted by the University of London for the M.Sc. degree.

[‡] Now at the British Caribbean Meteorological Service, Port-of-Spain, Trinidad.

§ 2. Existence and Properties of the Electrostatic Centre

We consider an insulated, uncharged conductor of arbitrary shape, *in vacuo* or in a homogeneous fluid dielectric of dielectric constant ϵ. If a uniform field is applied, the total potential ψ can be expressed in terms of the potential ϕ of the uniform field alone and the surface density σ of the induced charge on the conductor by

$$\psi=\phi+\int\frac{\sigma}{\epsilon r}\,dS, \qquad \ldots\ldots(1)$$

where r is the distance from the surface-element dS to the point at which ψ is taken. We are neglecting all effects such as electrostriction, non-linearity of the dielectric and a double layer, for while they are invariably present, they are usually small.

Suppose that the applied field is of strength $\mathbf{F}_1$ and is in the x direction. The potential due to this field alone may be given by

$$\phi_1=-F_1x \qquad \ldots\ldots(2)$$

and the total potential is the sum of this and a potential due to an induced density σ_1 according to (1). The potential ψ_{1c} acquired by the conductor will be equal to the value of ϕ_1 at some plane $x=p_1$. Suppose now that the strength of the applied field is changed to $k_1\mathbf{F}_1$, where k_1 is any constant. Then since $\mathbf{F}_1$ and the field due to σ_1 produce equilibrium on the surface of the conductor, so do $k_1\mathbf{F}_1$ and the field due to a density $k_1\sigma_1$. However, we know from the uniqueness theorem for charge distributions (Jeans 1943) that there can only be one equilibrium distribution for a given applied field. Thus, by necessity, when $\mathbf{F}_1$ changes to $k_1\mathbf{F}_1$, σ_1 changes to $k_1\sigma_1$, and it follows from (1) and (2) that the potential of the conductor is still the same as that of the uniform field alone at the plane $x=p_1$.

Similar arguments hold if the uniform field is taken to be in the y direction and of strength $\mathbf{F}_2$, or in the z direction and of strength $\mathbf{F}_3$. In the former case the potential ψ_{2c} of the conductor is the same as ϕ_2 of the applied field at a plane $y=p_2$ and in the latter case, ψ_{3c} is the same as ϕ_3 in a plane $z=p_3$, and these planes are the same when the fields are varied. We now consider the effect of applying $\mathbf{F}_1$, $\mathbf{F}_2$ and $\mathbf{F}_3$ simultaneously, so that the total applied field is uniform and of strength

$$\mathbf{F}=\mathbf{F}_1+\mathbf{F}_2+\mathbf{F}_3. \qquad \ldots\ldots(3)$$

It follows from the theorem of superposition (Jeans 1943) that on the conductor, the induced density $\sigma=\sigma_1+\sigma_2+\sigma_3$ and the potential $\psi_c=\psi_{1c}+\psi_{2c}+\psi_{3c}$. We therefore see that ψ_c is the same as the potential due to $\mathbf{F}$ alone at the point of intersection P of the planes $x=p_1$, $y=p_2$, $z=p_3$. We also see that P is the same when $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}_3$ are varied by arbitrary constant factors k_1, k_2 and k_3. But a change such that $(k_1F_1)^2+(k_2F_2)^2+(k_3F_3)^2=F_1^2+F_2^2+F_3^2$ is simply a rotation of $\mathbf{F}$ relative to the body, which is electrostatically equivalent to keeping $\mathbf{F}$ fixed and rotating the body about axes through the origin. For the latter rotation P may move, but it remains fixed relative to the body. Now any displacement of the body at all can be accomplished by a rotation about axes through the origin and a translation. The change of potential in a translation, in which a point on the body at N, say goes to N′, is given by

$$\psi_{cN'}-\psi_{cN}=\phi_{N'}-\phi_{N}, \qquad \ldots\ldots(4)$$

as follows from (1), so that P is translated with the body. We conclude that the point P is always fixed relative to the conductor. It is therefore the electrostatic centre according to the definition given.

There can only be one electrostatic centre, for if there were two, the potential of the conductor would have to be equal to that of the applied field at both of them, which would only be possible for certain orientations of the body.

Suppose that, instead of having zero total charge, the conductor has a charge $\overline{Q}$ and a corresponding potential $\overline{\psi}_c$. If the field $\mathbf{F}$ is applied, we have by the theorem of superposition that the total potential of the conductor is $\psi_c+\overline{\psi}_c$. It follows that the potential of the conductor minus $\overline{\psi}_c$ is always equal to the potential due to $\mathbf{F}$ at P, and with this generalization, our theorem on the electrostatic centre includes conductors which are not insulated or uncharged.

Associated with a total charge $\overline{Q}$ there is a density $\overline{\sigma}$. A further consequence of the theorem of superposition is that when the field $\mathbf{F}$ is applied, the total density is $\sigma+\overline{\sigma}$ and while σ depends upon $\mathbf{F}$, $\overline{\sigma}$ does not and represents a distribution fixed relative to the conductor. If $\overline{Q}$ is zero, the only mechanical action on the conductor is due to $\mathbf{F}$ upon σ, and is a couple, but if $\overline{Q}$ is non-zero, there is this couple together with a force. Consider a small rotation $d\theta$ of the conductor about an axis through P. Since $\overline{\sigma}$ moves with the body, the work done on the body on account of $\overline{\sigma}$ is equal to the work done on the distribution $\overline{\sigma}$ itself, which is zero because $\psi_c+\overline{\psi}_c$ remains constant. In terms of the total moment $\overline{G}$ of the forces upon the body due to $\overline{\sigma}$, the work is $\overline{G}d\theta$. Thus $\overline{G}=0$, and since this is true for any axis through P, the force due to $\overline{Q}$ acts through P.

§ 3. EXAMPLE

It has been mentioned before that the general existence of an electrostatic centre was first suspected on account of results for a special problem, treated by Vickers (1954). We shall discuss this problem very briefly and quote the relevant results.

We attempt to calculate the total electrostatic field when a uniform field of strength $\mathbf{F}$ influences a conductor the equation of whose surface is

$$r_1+\lambda r_2=d, \qquad \ldots\ldots(5)$$

where r_1 and r_2 are bipolar coordinates drawn from poles O_1 and O_2 say, and λ and d are constants. When $\lambda\neq1$, this surface is the surface of revolution of a Cartesian oval about an axis through the poles; it reduces to a prolate spheroid in the special case $\lambda=1$. We take the mid-point of O_1O_2 as pole and the line through these points as initial line of spherical polar coordinates (r, θ, ω). Equation (5) may be transformed to these coordinates and becomes a polynomial in r, $\cos\theta$ and a, the half-distance between O_1 and O_2.

The surface (5), which we shall refer to as an oval, is interesting from the electrostatic point of view essentially because it is asymmetrical, that is, it has no centre of symmetry, except when $\lambda=1$. There appears to be no way of obtaining an exact solution of the problem, but if a is small, the oval is quasi-spherical, and an approximate solution is obtainable by the use of spherical harmonics according to a well-known method (Jeans 1943). Assuming then that a is small, we may approximate to the spherical-polar form of (5) by expressing r as a polynomial in powers of a, the highest power determining the order of the approximation, which must be followed consistently throughout the calculation. To the third order, we have, in terms of Legendre polynomials,

$$r = \frac{d}{\lambda+1} - a\left(\frac{\lambda-1}{\lambda+1}\right)\mathrm{P}_1(\cos\theta) - a^2\frac{(\lambda+1)}{3d}\{1-\mathrm{P}_2(\cos\theta)\}, \qquad \ldots\ldots(6)$$

the coefficient of a^3 being zero.

Taking the pole and initial line as origin and x axis, respectively, of Cartesian coordinates, we choose the y axis to be in the plane determined by the x axis and $\mathbf{F}$. The potential due to $\mathbf{F}$ may thus be written

$$\begin{aligned}\phi &= -F(lx+my),\\ &= -Fr\{l\mathrm{P}_1(\cos\theta) + m\mathrm{P}_1{}^1(\cos\theta)\cos\omega\},\end{aligned} \qquad \ldots\ldots(7)$$

where l and m are the direction cosines of $\mathbf{F}$. We specify that any medium in which the oval may be immersed is a homogeneous insulating fluid of dielectric constant ϵ, so that the total potential ψ satisfies Laplace's equation. We then assume, in accordance with the theory of spherical harmonics,

$$\psi = \phi + \sum_{n=0}^{\infty}\frac{1}{r^{n+1}}\sum_{m=0}^{n}\mathrm{P}_n{}^m(\cos\theta)(A_{nm}\cos m\omega + B_{nm}\sin m\omega). \qquad \ldots\ldots(8)$$

We assume further, in keeping with the special method of handling problems of this type, that the coefficients are of the form

$$A_{nm} = \sum_{q=n}^{\infty} f_{qm}\,a^{\alpha+q}, \quad B_{nm} = \sum_{q=n}^{\infty} g_{qm}\,a^{\beta+q}, \qquad \ldots\ldots(9)$$

where α and β are constants and f_{qm} and g_{qm} are functions of λ and d. The boundary condition is

$$\psi = \text{const.} \qquad \ldots\ldots(10)$$

at the surface of the oval. The equation of the oval is used in the approximate form (6), so that the calculation of α and β determines a value of n at which the series (8) may be cut off consistently with this approximation. Thereafter, the calculation of the A_{nm} and B_{nm} is free of mathematical difficulties but is very laborious. The B_{nm} are found to be all zero, as are the A_{nm} for $m>1$, while A_{n0} and A_{n1} for $n=0, \ldots 4$ are non-zero and are of the correct form (9). It is found that A_{00} is proportional to the total charge on the oval, and that if $A_{00}=0$, the potential of the body is the same as that of the applied field at the point $(x', 0, 0)$, where

$$x' = -a\left(\frac{\lambda-1}{\lambda+1}\right)\left(1+\frac{8}{5}\frac{\lambda a^2}{d^2}\right). \qquad \ldots\ldots(11)$$

This point is dependent only upon the geometry of the oval and, to the order a^3, is the electrostatic centre.

As a matter of interest, we may note that the centre of gravity of the oval, regarded as a uniform solid, is the point $(x'', 0, 0)$, where

$$x'' = -a\left(\frac{\lambda-1}{\lambda+1}\right)\left(1+\frac{\lambda a^2}{d^2}\right). \qquad \ldots\ldots(12)$$

It will be observed that both the points (11) and (12) become the origin of coordinates when $\lambda=1$, which is as it should be, for the oval is then a prolate spheroid, for which both the electrostatic centre and the centre of gravity are at the geometrical centre.

When the potential ψ is known, the normal component of field strength at the surface of the oval, E_ν, may be calculated. If $\boldsymbol{\rho}$ and $\boldsymbol{\nu}$ are respectively the displacement vector from a fixed point N to a surface element dS, and the unit outward normal at dS, the torque on the oval about N is given by

$$\mathbf{G} = \frac{\epsilon}{8\pi}\int E_r^2[\boldsymbol{\rho}, \boldsymbol{\nu}]\, dS. \qquad \ldots\ldots(13)$$

The integration, to the order a^3, is elementary but tedious. It is found that if N is chosen to be the electrostatic centre, (11), $\mathbf{G}$ is independent of A_{00} and therefore of the total charge, showing that the force due to the total charge acts through the electrostatic centre.

ACKNOWLEDGMENT

We are grateful to Dr. F. H. Bowen for his support of this work, especially for extending to one of us (D.O.V.) facilities for post-graduate study.

REFERENCES

JEANS, J. H., 1943, *Electricity and Magnetism* (Cambridge : University Press).
VICKERS, D. O., 1954, *M.Sc. Thesis*, University of London.

An X-ray Study in High Vacuum of the Structure of Evaporated Copper Films

By H. D. KEITH
H. H. Wills Physical Laboratory, University of Bristol

MS. received 20*th June* 1955 *and in amended form* 17*th August* 1955

Abstract. Films of metallic copper which have been deposited on a glass substrate at − 183° c in a high vacuum to a thickness of 2000 Å are shown to consist of aggregates of randomly oriented micro-crystallites of nominal size 40 Å, this dimension being reproducible and almost independent of the nature of the substrate at this temperature. If the films are annealed to room temperature in a vacuum which precludes the chemisorption of oxygen or other reactive gas, there is a large, rapid and irreversible increase in the nominal particle size to at least 400 Å. The chemisorption of a saturated monolayer of oxygen at − 183 °c before the annealing takes place, however, almost completely prevents this re-crystallization and the particles grow to only 60 Å. The corresponding chemisorption of a smaller quantity of oxygen, sufficient only to give a small but appreciable fraction of a monolayer, inhibits the annealing process and can prolong it for several days. It does not, however, prevent the film from re-crystallizing to almost the same extent as in the complete absence of oxygen. These results are shown to be in qualitative agreement with resistivity and gas adsorption data for copper films, and their bearing on the problem of establishing a model for the structure and behaviour of evaporated metallic films is outlined. The experiments as a whole indicate that surface diffusion plays a leading part in the re-crystallization of the disordered films.

Some further experiments on the effect of residual air on the structure of copper films deposited at room temperature are described, and the design of a low temperature high vacuum x-ray focusing camera, which was constructed entirely of glass and used for the above experiments, is also discussed.

---

§ 1. INTRODUCTION

It is well known that thin metallic layers which have been deposited on cold substrates from atomic beams *in vacuo* have properties substantially and occasionally markedly different from those of the corresponding bulk metals. These anomalous properties are believed to be due to very small particle size and to relatively large concentrations of lattice imperfections. Both factors depend on a number of variables but the most important of these appears to be the relation between the temperature of the substrate before deposition and the melting point of the metal deposited. The work of Buckel (1954) and of Rühl (1954) indicates that gallium, bismuth and tin containing about 10% of copper show little vestige of their normal crystallinity when deposited at temperatures between 2° K and 20° K. They undergo sharp transitions to a disordered crystalline state, however, at about 20° K and it is probable that most, if not all, metals would behave similarly if deposited at sufficiently low temperatures. As the deposition temperature (or the temperature of an already deposited film) is raised, the size of the crystallites constituting the film increases. Thus, at room temperature, films of metals having low melting points re-crystallize and consist of fairly large crystallites, whereas with metals having higher melting points the crystallites remain small (Levinstein 1949, Umanskii and Krylov 1946). An increase in the crystallite size during re-crystallization is accompanied by the annealing out of lattice imperfections. Vand (1943) has attempted to describe the anomalous properties of thin films wholly in terms of lattice imperfections formed during deposition, each of which is assumed to become mobile when the temperature appropriate to its characteristic activation energy is attained. This approach, however, is not entirely satisfactory as is evident from the efforts made by Wilkinson and Birks (1949) to determine and interpret the spectrum of activation energies of the imperfections in films of gold.

Much remains to be learnt about the structure of thin metallic films and particularly about the mechanisms involved in their re-crystallization. The experiments described in this paper were carried out in order to measure the sizes of the constituent crystallites in evaporated copper films, both at liquid oxygen and at room temperature. The effect of chemisorbed gaseous layers on the annealing of the films was also studied since it was thought that the observed re-crystallization was probably due in large measure to the diffusion of atoms over the surfaces of the small crystallites. Copper was chosen because other properties of evaporated films of this metal have been made the subject of exhaustive researches in this laboratory by Dr. J. W. Mitchell and his students and the author is indebted to these workers for access to their results before publication.

The measurements of Mitchell and his collaborators show that copper films, when deposited on glass at -183° c to a thickness of 500–1000 Å, have resistivities roughly 30 times greater than bulk copper. After deposition there is a small spontaneous decay of resistivity following the hyperbolic law (Vand 1936); and the resistivity then falls irreversibly to about three times the bulk value if the films are annealed to room temperature in conditions where gaseous or other contamination is avoided. The results of exposing the films to oxygen at -183° c before annealing depend on the quantity of gas admitted. With large amounts the resistivity rises sharply and falls during subsequent annealing, but to a value still greater than that initially found at the lower temperature. With smaller quantities of oxygen, the resistivity falls on annealing to considerably below its initial value but the admission of oxygen always gives a larger final resistivity at room temperature than would otherwise be found. Corresponding measurements of the adsorption of gases by films deposited at -183° c indicate that the actual surface areas of these films are about ten times greater than the areas calculated geometrically, whereas in the case of annealed films they agree much more closely. The densities of the films were deduced from measurements of mass and thickness, and at both temperatures were 15 to 20% below the value for bulk copper. In spite of the diversity and reproducibility of these results, and those of many other workers, the difficulty of building up a detailed model of the structure of the copper film is still great, and the present investigation was undertaken in the hope that x-ray diffraction studies might provide useful evidence bearing on the problem.

The correlation between the results described above and those obtained from the diffraction experiments is discussed in § 4.

§ 2. THE X-RAY DIFFRACTION CAMERA AND ITS OPERATION

Many designs for vacuum x-ray powder cameras have appeared in the literature (see for example Rühl 1954), but no attempt appears to have been made hitherto to achieve operating vacua comparable with those obtainable in sealed-off and gettered glass envelopes. Though seldom necessary, this is clearly essential for an investigation in which a metallic surface has to be maintained free from gaseous contamination for many hours, and the camera now to be described is, in effect, a glass vacuum tube.

The camera was designed to record a number of Debye–Scherrer diffraction lines in a form suitable for examining line profiles and particle size broadening. The latter effect depends on the function $\sec\theta$ where θ is the Bragg angle, but the sensitivity of measurement is only high for a broadening of moderate extent. Hence it follows that, if during an annealing process the particle size changes by an order of magnitude, it is necessary to study more than one diffraction line, observing a line at a small value of θ before and one at a larger value of θ after annealing. The (200) and (311) copper lines ($\theta = 22{\cdot}5^\circ$ and $45{\cdot}0^\circ$ respectively with Cu $K\alpha$ radiation) were, in fact, chosen. In order that a reasonable diffracted x-ray intensity could be obtained from copper films only 2000 Å thick without using a wide entrance slit, a focusing camera of the Seeman–Bohlin type was constructed. This enabled a considerable area of evaporated film ($\sim 3\ \text{cm}^2$) to be employed and the instrumental broadening to be kept small.

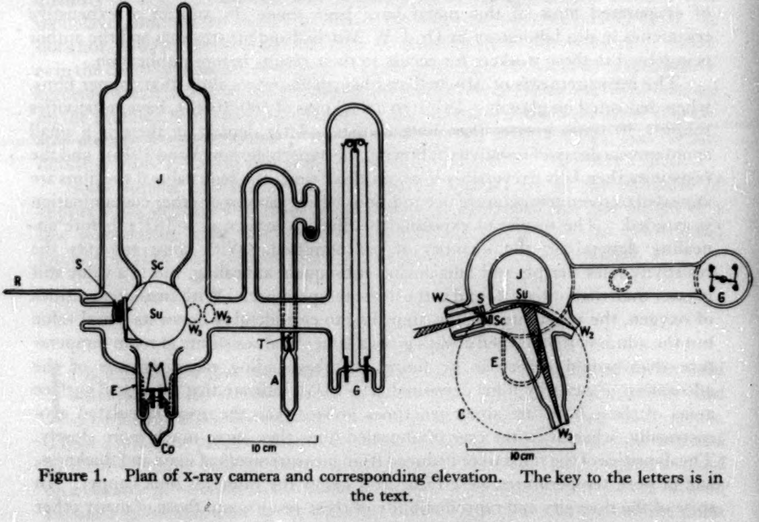


Figure 1. Plan of x-ray camera and corresponding elevation. The key to the letters is in the text.

The glass substrate on which the films were deposited was made part of the wall of a receptacle for liquid oxygen. Thus the camera, which is illustrated in figure 1, resembles a Dewar vessel with the cylindrical substrate Su attached to the lower end of the inner jacket J. The various Bragg reflections can be seen in the plan view to converge to specific positions on the cylindrical focusing surface C; and the two strong reflections chosen for study fall on the windows W_2 and W_3. By locating these and the entrance window W_1 as shown, the slight low angle scattering caused by them can have no effect on photographic films supported in air against W_2 and W_3. The windows are made of very thin blown borosilicate glass, are capable of supporting a vacuum even at 400° c and transmit about 70% of Cu Kα radiation falling normally upon them.

The methods used to execute the design in glass with adequate precision are possibly of sufficient general interest to be given a brief mention at this point. The cylindrical substrate was cut from the wall of a carefully chosen glass tube with the required internal diameter (12 cm) and free from eccentricity. The section was ground to have a circular projection on to the tangent plane at its centre and sealed to the end of a previously prepared and shaped glass tube. A T-piece was then blown close to this to provide the inner jacket of the camera. The most critical stage in the procedure is the location and alignment of the entrance slit S, which is defined by two carefully ground semi-circular pieces of glass held by individual glass supports from the inner jacket and platinized to improve their opacity to x-rays. In view of the proximity of the slit to the substrate, any small error in its location not only displaces the diffraction maximum but also contributes alarmingly to instrumental broadening. It can, however, be positioned accurately in the following way. The edges of the substrate are lightly ground to remove protrusions above the cylindrical profile due to glass blowing. A thin strip of sheet brass (1/32 inch thick) is then rolled to fit the inside surface of the tube from which the substrate was cut. If placed in intimate contact with the substrate this automatically defines the focusing surface and one slit member can be placed in position followed by the other, the slit width (~0·2 mm) and its uniformity being judged with a feeler gauge. Simple optical tests can be used to check the parallelism of the slit, which is 1 cm in length, with the cylinder axis. A small scatter shield Sc is used to prevent fogging of the film located at W_2. The positions of the windows W_2 and W_3 are calculated and their location achieved by cutting triangles from rigid molybdenum sheet such that when one apex is inserted in the slit, and a second allowed to rest on the substrate Su, the third determines the required focal point. It was possible to construct cameras which, despite the obliquity of the beam falling on the specimen (20°), are capable of clearly resolving the Cu Kα doublet at $\theta = 45°$ without any adjustment based on trial and error.

The cameras were prepared for evacuation taking the precautions customary in high vacuum work. Spectroscopically pure copper beads were fused in hydrogen on to molybdenum filaments and sealed into the evaporator tube E from which they could be deposited on the substrate with as little obliquity (~45°) as is permitted by the presence nearby of the arm carrying the window W_3. The distance of the beads from the centre of the substrate was 7 cm. In the side arm G molybdenum films can be deposited from central hairpin filaments to provide an extremely efficient getter for oxygen and nitrogen (Rivière 1954). Gas ampoules A, containing oxygen which had been prepared by heating potassium permanganate *in vacuo*, were added in some of the experiments and access to the tube effected, when required, by breaking the fine glass tip T with a glass-encased ball bearing operated magnetically. After several alternate bakings in an oven at 400° c and thorough outgassings of the metal parts by heating, the cameras were sealed off at constrictions Co provided for the purpose. The pressure at this stage was never greater than 10^{-7} mm Hg as assessed by means of an ionization gauge, and subsequent firing of the getter reduced the pressure in trial experiments to unobservably small values. Simple calculations based on the number of molecules present and the surface area of the molybdenum film show that all the residual reactive gas can easily be accommodated as a chemisorbed monolayer on the molybdenum. The copper films were not deposited until at least several hours later and there can be little doubt that the bulk of the deposit is free from surface contamination. Because of the time involved in making the cameras, they were generally used a large number of times, being thoroughly cleaned between runs and the pinches and getter tubes being replaced.

X-rays were generated in a conventional horizontal x-ray tube with a copper target and monochromatized by means of a curved quartz crystal. This produced a beam of Cu Kα radiation with an angular spread of 5° converging to a vertical line focus approximately 1 cm long. The only difficulty encountered was that of ensuring that the line focus fell on the entrance slit which should be completely flashed by the beam. The entire camera is held rigidly in a stand with a kinematically located base and the initial alignment carried out in the following way.

A glass rod R was sealed to the wall of the camera and arranged to run parallel to and above the incident x-ray beam. Two short lengths of straightened molybdenum wire, one 0·3 mm and the other 0·2 mm in diameter, were then suspended from plasticine attached to R at convenient distances, approximately 10 cm and 6 cm from the entrance slit. The adjustment of these wires and the slit for parallelism and collinearity was effected by illuminating the slit from behind and using the observed diffraction effects as a guide. When inserted in the x-ray beam the two wires cast shadows which are almost equally distinct on a fluorescent screen placed outside the camera and about 3 cm in front of the slit. A gentle rotation of the camera brings these shadows together. At one setting the combined shadow exhibits a marked lateral contraction, which affords a critical test for the alignment. This rather crude method was initially a make-shift, but proved so simple and successful that it was retained.

The instrumental breadth at half intensity of the camera for the low-angle (200) reflection was 0·59° which is close to the theoretical value for the slit width used (0·18 mm). Unfortunately, the monochromator gives a partial separation (~0·1 mm) of the two components of the Cu Kα doublet so that a slight alteration in camera setting changes the ratio of the intensities of the components received by the specimen. Where broad reflections were being studied it was convenient to arrange that little or no α_2 radiation passed the slit, whereas with well annealed films, when the doublet could be resolved, the α_1 and α_2 radiations were allowed to pass in roughly equal proportions. This enabled a sensitive visual assessment of resolution to be made which is invaluable in cases where the broadening is just detectable. The complication introduced by the doublet structure of the radiation only becomes appreciable with the (311) reflections, but it limits the accuracy of particle size measurements very severely when the particles are large.

Exposures of 120–160 mA hours at 35 kV were used for annealed copper films and up to 240 mA hours for highly disordered films. The diffraction lines were recorded on Ilford P.M.1 film which had sufficient speed and contrast and was obtainable in conveniently small packs.

§ 3. EXPERIMENTAL RESULTS

All the copper films were deposited to an estimated thickness of 2000 Å at very nearly the same rate of 10 Å per second throughout. Whatever the deposition temperature, the films always gave uniform diffraction lines which were of the expected relative intensities, and no evidence of preferred orientation was found. A preliminary investigation of similar copper films which was carried out in air with a conventional Debye–Scherrer camera also failed to reveal any preferred orientation. The diffraction lines were always free from spottiness, showing that large islands of single crystal were not formed.

The nominal particle sizes quoted below were estimated from the broadening of the diffraction lines as measured with a microdensitometer, but a word of explanation must be said in this connection. Owing to the limitations imposed by the restricted aperture of the windows of the camera and the broad diffuse tails of some of the diffraction lines, half-intensity line breadths rather than integral breadths were measured. This method is more reliable for these experiments and the correction for instrumental breadth is easily applied. The formula $B = K\lambda \sec\theta / t$ which relates the corrected angular broadening B of the diffraction line at half maximum intensity with the effective mean crystal diameter t normal to the reflecting planes considered, is seen to involve a constant K for which various values (depending on the shape of the crystals) have been given, all of which are roughly unity. It was decided to adopt the value $K = 1{\cdot}0$ throughout and to refer to the results so obtained as 'nominal particle sizes'. There seemed to be little point in applying more refined methods in view of the difficulty in distinguishing between the effects of size distribution and shape distribution of the particles. A discussion of the question as to whether alternative interpretations of the considerably broadened diffraction line profiles are possible or likely is deferred until § 4.

3.1. *Films Deposited and Maintained at* − 183° C *in High Vacuum*

The films deposited at − 183° C had, in all cases, a dark murky green appearance and gave diffraction lines which were extremely broad and diffuse (see figure 2(*a*) Plate). In fact, the (311) line was not detected at all and a faint but unmistakable trace of a very broad (200) line was only found with prolonged exposures.

The nominal particle size for copper films in this condition is 40 Å, a figure which was found reproducibly within an experimental error of ± 5 Å. Since the exposures were much longer than the time required for the spontaneous decay of resistivity in such films there was no possibility of detecting related effects in the diffraction patterns.

3.2. *Films Deposited at* − 183° C *and Annealed to Room Temperature in vacuo*

When the murky green films described in the previous paragraph are allowed to warm to room temperature they rapidly assume the normal characteristic colour of metallic copper. This colour change appears to be complete within a few minutes. X-ray exposures were in some cases commenced immediately after this annealing and in all cases resulted in photographs such as that shown in figure 2 (*b*). The nominal particle size is now very difficult to judge for the reasons given earlier since the particles are at least 400 Å in size and may be much larger. Cooling the films again to − 183° C, produces no detectable change in the line width. Annealing the films to 400° C for two hours results in little or no further reduction in line width (figure 2(*c*)), but the diffuse tails of the line profiles disappear (see figure 3). These diffuse tails also diminish in intensity if the films are left standing at room temperature for several days.

3.3. *Films Deposited in vacuo at* −183° c *and Annealed in Oxygen*

On the basis of the measurements of Allen *et al.* and from the geometry of the camera, getter tubes and deposited films, it is possible to estimate with fair accuracy the effective surface areas of the copper film at liquid oxygen temperature and the molybdenum film at room temperature. Assuming that there are about $1{\cdot}5 \times 10^{-15}$† adsorption sites per cm², the number of oxygen molecules required to cover the surface completely as a chemisorbed monolayer can be calculated to within a factor probably not greater than 5 and almost certainly less than 10. In one series of experiments gas ampoules were filled with oxygen to a pressure which, when expanded in the camera, was sufficient to ensure at least two monolayers even if the uncertainties and errors in the calculation accumulated unfavourably. All the oxygen was assumed to have a chance of being chemisorbed by the metal, there being good evidence to show that the adsorptive power of borosilicate glass is small at −183° c and negligible at room temperature (Dorling 1953).

Films annealed to room temperature after being exposed to oxygen in this way for several minutes at −183° c gave diffraction lines as shown in figure 2(*d*) and the nominal particle size is now only 60 Å. This result was also reproducible. The oxygenated films retained their murky green appearance and were quite stable at room temperature. Exposures taken after they had been standing untouched for four weeks still gave diffraction lines indistinguishable from those just mentioned. On annealing to 200° c for two hours without further admission of gas, the murky green appearance gives way slightly to a faint red metallic sheen and the nominal particle size increases to 200 Å.

The results of these experiments at once suggest the admission of oxygen to the camera at a pressure, after expansion, of $1{\cdot}5 \times 10^{-1}$ mm Hg to serve as a 'brake' on the annealing, thus providing a very useful experimental tool for freezing-in transient conditions of the films for later and more leisured study. This artifice was used in some of the experiments to be described below, but has not yet been fully exploited.

Films were also annealed after being exposed at −183° c to oxygen at one hundredth of the pressure used in the experiments just described. The films so treated retained their murky green appearance for some time after annealing to room temperature, and it was only after 12 hours that a slight colour change could be recognized with confidence. This trend continued steadily and 72 hours had elapsed before a genuine copper colour became well developed. The result of an x-ray exposure commenced after this stage had been reached is shown in figure 2(*e*) and the line broadening is sufficient only to prevent resolution of the doublet structure of the (311) line. If a second ampoule containing enough oxygen to act as a 'brake' is opened 24 hours after the initial anneal in another similar experiment, the photograph shown in figure 2(*f*) is obtained. This shows that although the effect of the adsorption on the colour of the annealed film is marked, the initial increase in particle size is not greatly retarded. It is inferred that the retardation in growth is only pronounced once the particles become large and the surface tension forces less effective; the colour change also appears to be associated with the later stages of the anneal.

---

† This figure is obtained by averaging the probable values for the (100), (110) and (111) faces.

3.4. *Films Deposited at* −183° c *on top of Thin Annealed Copper Films*

These experiments were carried out with the object of establishing whether the small size of the particles deposited at −183° c was determined by nucleation considerations or by the vitreous nature of the glass surface. A thin layer of copper, sufficient only to constitute a clearly visible and continuous deposit, was evaporated at −183° c and annealed to room temperature. This film appeared green in transmission, but reflected light of the same colour as thicker annealed films. It was then cooled to −183° c and the remaining copper deposited. For convenience, oxygen was admitted as a 'brake' and the completed film warmed to room temperature, when it retained its murky appearance. The diffraction photographs, figure 2 (*g*), show that the nominal particle size is now 100 Å, which is to be compared with 60 Å when the films are treated similarly after direct deposition on glass at −183° c. Due to extinction the thin annealed portion of the film has little effect on the line profile. Heating the film to 200° c for two hours reduces the broadening but does not entirely remove it. The conclusion to be drawn from this is that the particle size is determined principally by the temperature of the substrate and the nucleation of the crystallites, and not by the surface structure of the substrate.

3.5. *Films Deposited at Room Temperature in High Vacuum*

From the foregoing results it was natural to expect that films deposited at room temperature would have large nominal particle sizes and this was indeed found to be the case (see figure 2 (*h*)). For the reasons discussed previously, it was difficult to say with certainty whether these sizes are different from those found with films which have been annealed from −183° c to room temperature in vacuum. Visual observation of the doublet (311) line suggests, however, a slightly larger size than was found in the latter case even after annealing to 400° c, a result which is possibly connected with a smaller number of effective nuclei for re-crystallization. The photograph 2 (*h*) also shows to best advantage the resolving power of the camera for the (311) reflection at $\theta = 45°$.

3.6. *Films Deposited at Room Temperature in Air at Low Pressure*

In this case the effect of incorporating small quantities of air into the films during deposition was investigated. When a residual air pressure of 10^{-5} mm was maintained during deposition and reduced to below 10^{-6} mm immediately afterwards for sealing off, the films gave diffraction lines not appreciably different from those described in the previous paragraph. Films deposited in a similar manner with a residual air pressure of 10^{-4} mm did exhibit slight line broadening which disappeared when re-examined after 48 hours, a result clearly connected with the effects observed when films were annealed in the presence of small amounts of oxygen.

3.7. *Films Deposited at Room Temperature in High Vacuum and Stabilized by admitting Air*

Some experiments were conducted in which films were deposited at room temperature in an air pressure of 10^{-7} mm and air admitted at atmospheric pressure to the camera for 20 seconds immediately after the deposition ceased. Photographs such as that shown in figure 2 (*i*) resulted and did not change subsequently with time. The broadening so revealed, which would have annealed out fairly rapidly in vacuum, is presumably related to the disorder whose disappearance accounts for the spontaneous decay of resistivity.

These experiments clearly demonstrate the limitations of techniques in which, for example films are deposited in vacuum and removed almost immediately for diffraction experiments in air. Rühl (1954) reports that the growth texture of tin films depends on the quality of the vacuum in which they are deposited.

§ 4. Discussion of the Results

The experiments described in the previous section do not permit a detailed model of the copper films to be advanced with confidence, but the importance of surface diffusion in the re-crystallization of the films is believed to have been established. This discussion will therefore be brief and confined to three main topics: (*a*) the reasons for believing that almost all the observed line broadening is due to small particle size, (*b*) the qualitative correlation between the results of the diffraction experiments and other known properties of copper films, and (*c*) possible mechanisms for the re-crystallization of the films. Some general observations and suggestions for further experimental investigations will also be outlined.

(*a*) *Interpretation of the observed line broadening.*

Broad x-ray diffraction lines can be due to either (i) small particle size, (ii) lattice strain, or (iii) certain types of lattice imperfection. Of these, (i) requires little comment, but it should be pointed out that the formation of twins or stacking faults in the small crystallites would broaden diffraction lines substantially and result in fictitiously small values for the nominal particle size. However, from the fact that the peaks of the (200) diffraction lines obtained with films deposited and maintained at −183° c are not displaced perceptibly from their normal positions for annealed films, it can be concluded (Paterson 1952, Warren and Warekois 1953) that the frequency of occurrence of stacking faults is small and few crystallites of the observed size will be affected; furthermore, twinning is generally found only in oriented copper films deposited on crystalline substrates or as the result of annealing at elevated temperatures and is not observed in the cast metal. These complications are therefore thought to be unimportant in our experiments.

Under (ii) above, we shall consider curvature and distortion of the lattice due to microstresses and dislocation networks generally; certain other effects which might be considered under this heading are discussed later. Strain causes broadening of diffraction lines in Debye–Scherrer photographs by varying the dimensions of the unit cells in different parts of the crystal. The broadening of the (200) reflections observed with copper films deposited and maintained at −183° c was found within the limits of experimental error to be symmetrical and of a magnitude which, if interpreted as being entirely due to lattice strain, would correspond to a mean square strain of approximately 0·04. The largest mean square strain observed with severely cold-worked metals or alloys is about 0·008 or less (Warren and Warekois 1953, Haworth 1937), and greater strains than this are generally relieved spontaneously by plastic deformation. Comparison of these figures, the symmetrical nature of the broadening and the results of annealing films in oxygen allow this possibility to be ruled out, and it is concluded that at most only a small fraction of the observed broadening is due to strain. The fact that the films are deposited atom by atom in a disordered state and not subjected to harsh mechanical treatment must be borne in mind, but the existence of tensile stress in the films (Murbach and Wilman 1953, Hoffman, Daniels and Crittenden 1954) cannot be overlooked.

The alternative (iii) requires careful attention. Let us examine the not unlikely possibility that an appreciable concentration of vacant lattice sites is built into the films during deposition. The known density enables an upper limit to this concentration to be specified as approximately 15%. These holes will contribute to a diffuse background scattering of x-rays and may alter the mean lattice spacing. It is easily shown from Babinet's principle, however, that a random array of holes would give rise to a very diffuse background, but cannot contribute noticeably to the breadth of diffraction lines of the normal lattice. Chains of holes in $\langle 110 \rangle$ directions, a not unlikely configuration if the holes are mobile, could easily produce broadening of the (200) reflection, as is evident from a representation in reciprocal space; they could not, however, affect the (311) reflection and, since both reflections are in fact broadened proportionately, this possibility is also ruled out. If the distribution of holes is not random then some broadening of the line profiles may be expected but in view of the limitation on the integrated intensity of the diffraction due to the disorder, this will not be very significant. Some of the vacancies may migrate to internal surfaces or form planar aggregates but again little effect is to be expected beyond that which can be genuinely attributed to particle size or a moderate amount of strain.

Allowing for experimental error, the maximum possible alteration in the mean lattice spacing is found to be 0·2%, but whether or not this indicates a small vacancy content is not clear since the degree of relaxation of the atoms surrounding a vacancy in copper is not known accurately (Huntingdon 1942).

It is thought that the alternative (i) is the correct one, but whether the films consist of small individual crystallites or of larger crystals broken up by dislocation arrays into smaller blocks which scatter incoherently with one another is uncertain. The latter is very unlikely since, if it were the correct interpretation, the results of annealing the films in oxygen would have to be explained by assuming that oxygen can penetrate readily into the sub-boundaries at −183° c. The films are therefore believed to consist of aggregates of small crystallites which may themselves contain lattice disorder, but from the density of the films and from the marked influence of small quantities of oxygen most of the space responsible for the anomalous density would seem to be of larger than atomic dimensions. Here one has the elegant experiments of Hilsch (1954) on aggregating model atoms as a useful guide. An examination of the line profiles indicates that, although the half intensity breadth decreases markedly when a film is annealed *in vacuo* from −183° c to room temperature, the angular spread of the diffuse tails of the line is not greatly decreased. Much of the intensity originally diffracted into the region of the tails is now to be found, however, within the sharp central peak. After annealing to 400° c for two hours the tails disappear and the line profiles assume a more normal form. These changes are shown in figure 3. It is concluded therefore that in the anneal to room temperature either some of the crystallites remain fairly small or some of the larger re-crystallized grains are rod- or plate-shaped and randomly oriented. Further annealing to a higher temperature completes the re-crystallization and probably induces some grain growth.

(*b*) *The qualitative correlation between the results of the diffraction experiments and previous work.*

From the foregoing experiments, it appears that copper films deposited at −183° c consist of aggregates of randomly oriented microcrystallites whose size is about 40 Å on a glass substrate and about 50% greater on an annealed copper substrate. The large resistivities and surface areas of these films are associated with this structure, the latter arising naturally and the former as a combination of contact resistance and electron scattering at imperfections and crystal surfaces. The fact that there is little change in going from a vitreous substrate to the ideal case of relatively large copper crystallites clearly indicates that the structure of the films deposited at low temperatures is primarily determined by the nucleation of groups of condensing atoms. The nucleation of a fresh crystallite is perhaps the result of a fluctuation in the atomic beam density, and careful experiments in which deposition rates are widely varied without altering the heating of the films, by interrupted evaporation for example, would be most enlightening in this connection.

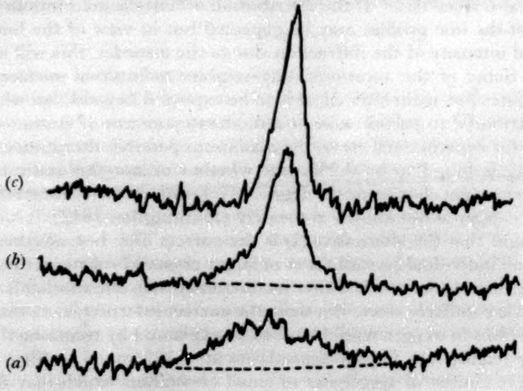


Figure 3. Microdensitometer records for the diffraction lines shown in figure 2 (*a*), (*b*) and (*c*).

The spontaneous decays of resistance observed by other workers are associated with the annealing out of those imperfections which provide the largest intrinsic contribution to the free energy of the disordered film, and whose gradual removal can probably be identified with the relaxation process invoked by Hoffman *et al.* (1954) to account for the observed tensile stress in metallic films.

The result of annealing uncontaminated films is to increase the particle size by at least an order of magnitude, thereby decreasing the anomalous resistivity, and at the same time diminishing the effective surface area. The effect on resistivity of annealing the films in oxygen at various pressures must be interpreted as being due to (i) the resistivity decreasing with increasing particle size and (ii) the contact resistance increasing when oxygen is adsorbed at or around the contact area.

The results of the experiments carried out with films deposited at room temperature will not be considered further, their place in the general scheme being obvious from the foregoing discussion.

(*c*) *Possible mechanisms for the re-crystallization of the films.*

We have seen that there appears to be no satisfactory explanation of the experimental results other than one in which the uncontaminated crystallites are assumed to grow rapidly in size on annealing to room temperature. A small part of the line sharpening may be due to the removal of strain, and it is likely that the change in nominal particle size from 40 Å to 60 Å when the films are annealed in excess oxygen is due to this effect. It is difficult to see how a monolayer chemisorbed at a low temperature could prevent the annealing out of dislocations and other lattice imperfections, so that the maximum amount of lattice strain to be relieved must therefore be very small, as stated earlier.

The mechanisms responsible for re-crystallization are severely hindered by a chemisorbed layer of oxygen and the major contribution must therefore be from surface diffusion. Oxygen chemisorbed on copper cannot desorb at room temperature (Winter 1954) and a saturated monolayer would prevent the detachment and migration of surface copper atoms. An incomplete monolayer will clearly inhibit the process, but provided that sufficient channels remain among the chemisorbed oxygen 'molecules' for the movement of copper atoms, the films can still re-crystallize. However, the large observed increase in particle size is, perhaps surprising and the manner in which larger crystallites engulf smaller ones is not yet understood.

A possible process is one in which some crystallites at internal free surfaces are denuded layer by layer, the copper atoms migrating over the free surface to be incorporated in growing crystals which present surfaces on which, for crystallographic reasons, the conditions for growth are more favourable. Rod- or plate-like crystals could be formed in this way as an internal channel is filled up with material seeded on one or more of the crystallites bordering on it; and such crystals, if randomly oriented, would give diffraction line profiles of the type found. A suitable size distribution of more rounded particles, some of which could grow at the expense of others in a similar way, would also account for the observed profiles and at present there seems to be no simple way of resolving this ambiguity. A more detailed study of surface diffusion processes under controlled conditions, however, will doubtless enable further information to be obtained. The methods developed by Warren and Averbach (1953) and others for investigating in detail the structure of cold-worked metals could be applied profitably to thin films stabilized with oxygen for study in air or in an inert gas atmosphere.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the collaboration of Mr. J. H. Burrow to whose skill in glass-blowing the good performance of the camera is entirely due, and whose advice and experience contributed much to its design. The author is deeply indebted to Dr. J. W. Mitchell for permission to quote some of the results of his researches before publication, and also for most helpful discussions. Dr. P. M. B. Walker and Mr. B. M. Richards of King's College, London kindly offered the use of their recording microdensitometer for studying the x-ray photographs. Mr. K. H. Koch rendered valuable assistance with the experimental work.

REFERENCES

BUCKEL, W., 1954, *Z. Phys.*, **138,** 136.
DORLING, E. B., 1953, *Ph. D. Thesis*, University of Bristol.
HAWORTH, F. E., 1937, *Phys. Rev.*, **52,** 613.
HILSCH, R., 1954, *Z. Phys.*, **138,** 432.
HOFFMAN, R. W., DANIELS, R. D., and CRITTENDEN, E. C., 1954, *Proc. Phys. Soc.* B, **67,** 497.
HUNTINGTON, H. B., 1942, *Phys. Rev.*, **61,** 325.
LEVINSTEIN, H., 1949, *J. Appl. Phys.*, **20,** 306.
MURBACH, H. P., and WILMAN, H., 1953, *Proc. Phys. Soc.* B, **66,** 905.
PATERSON, M. S., 1952, *J. Appl. Phys.*, **23,** 805.
RIVIÈRE, J. C., 1954, *Ph.D. Thesis*, University of Bristol.
RÜHL, W., 1954, *Z. Phys.*, **138,** 121.
UMANSKII, M. M., and KRYLOV, V. A., 1936, *J. Exp. Theor. Phys.*, **6,** 691.
VAND, V., 1936, *Z. Phys.*, **104,** 48; 1943, *Proc. Phys. Soc.*, **55,** 222.
WARREN, B. E., and AVERBACH, B. L., 1953, *Modern Research Techniques in Physical Metallurgy* (Cleveland : American Society for Metals), p. 95.
WARREN, B. E., and WAREKOIS, E. P., 1953, *J. Appl. Phys.*, **24,** 951.
WILKINSON, P. G., and BIRKS, L. S., 1949, *J. Appl. Phys.*, **20,** 1168.
WINTER, E. R. S., 1954, *J. Chem. Soc.*, p. 3342.

The Propagation of Ultrasonics in Suspensions of Particles in a Liquid

By J. BUSBY and E. G. RICHARDSON
Physics Department, King's College, University of Durham, Newcastle-upon-Tyne

MS. received 11*th July* 1955 *and in amended form* 12*th September* 1955

Abstract. The propagation of sound in suspensions of solid particles in water in the range of frequency 1 to 10 Mc/s and of particle radius 8 to 100 microns is measured. For the ultrasonic attenuation a pulse method is used, and the suspension is either monodisperse (glass spheres) or polydisperse (silica particles). In the latter suspension the particle size distribution is measured optically.

Results are compared with a theory due to Sewell and others and good agreement obtained. The ultrasonic absorption is linear with concentration up to 18% by volume. Some measurements were also made on 'sediments' of glass spheres and natural sands in which the attenuation is high and increases rapidly with frequency and particle size.

§ 1. Introduction

THE first calculation of the attenuation α of the amplitude in a beam of sound due to the insertion therein of a single sphere (small compared with the wavelength) was made by Rayleigh (1878).

Laidler and Richardson (1938) investigated the propagation of sound in suspensions of *lycoperdon* (puff-ball) spores and magnesium oxide in air in the kilocycle region and found partial agreement between their measurements and the calculations from Sewell's (1910) theory, but Hartmann and Focke (1940) in suspensions of such spores in water at megacycle frequencies found some disagreement, apparently due to the movement of the spheres with the sound field, an effect negligible in aerosols but significant in hydrosols. Lamb (1916) calculated the velocity potential of waves scattered from a mobile sphere and, by integration over a surrounding volume, the energy lost in this fashion. Urick (1948) cast Lamb's theory in a more convenient form and obtained for the attenuation due to a single (small) sphere free to move in (unit volume of) a viscous fluid the formula

$$\alpha = \frac{2\pi}{9}k^4a^6 + \frac{2\pi}{3}k(\sigma-1)^2a^3\frac{S}{S^2+(\sigma+\tau)^2} \qquad \ldots\ldots(1)$$

where a is the radius of sphere, $k=2\pi/\lambda$, λ the wavelength of the radiation; $\sigma=\rho_1/\rho_2$ is the density of particles relative to fluid,

$$S=\frac{9}{4\beta a}\left(1+\frac{1}{\beta a}\right), \qquad \tau = 1+\frac{9}{4\beta a},$$

β is the propagation constant of shear waves in the fluid, i.e. $(\omega/2\nu)^{1/2}$ where ω is the pulsatance and ν the kinematic viscosity.

These calculations then give the increase in the attenuation of a sound beam passing through a viscous fluid due to the intrusion of a single rigid sphere free to move in the sound field. To the extent that the attenuation produced by one sphere is unaffected by its neighbours we may adapt the formulae of Lowan *et al.* to our case and write an equation for the attenuation in a suspension in the form

$$\alpha=\frac{\lambda^2}{2\pi}\left[\sin^2 F_0\gamma+3\left(\frac{\sigma-1}{\sigma+\frac{1}{2}}\right)^2\sin^2 F_1\gamma+\sum_{m=2}^{m=\infty}(2m+1)\sin^2 F_m\gamma\right]+\frac{2\pi k(\sigma-1)^2a^3S}{3\{S^2+(\sigma+\tau)^2\}} \qquad \ldots\ldots(3)$$

or

$$\frac{\alpha\lambda}{C}=\frac{3\pi}{\gamma^3}\left[\sin^2 F_0\gamma+3\left(\frac{\sigma-1}{\sigma+\frac{1}{2}}\right)^2\sin^2 F_1\gamma+\sum_{m=2}^{m=\infty}(2m+1)\sin^2 F_m\gamma\right]+\frac{\pi(\sigma-1)^2S}{S^2+(\sigma+\tau)^2} \qquad \ldots\ldots(4)$$

where C stands for the volume concentration of spheres (volume of solid/total volume). The relative magnitudes of the scattering and viscous losses are given in figure 5.

§ 2. APPARATUS

2.1. *Ultrasonic and Electronic System*

In the experimental investigation of this relationship a pulse technique was used.

Most previous investigators using a pulse system have used a single crystal as both transmitter and receiver but in this work a pair of matched quartz crystals was used, one to transmit and one to receive. This removed difficulties due to the transmitted pulse passing into the receiver circuit and causing saturation.

A block diagram of the apparatus is shown in figure 1. The transmitter section comprises an asymmetrical multivibrator and an electron-coupled Hartley oscillator generating pulses of r.f. voltage. The length of these pulses could be varied between 2 and 25 msec. The receiver section consisted of (i) a network to match the 75 ohm input impedance of the attenuator to the impedance of the crystal, (ii) an attenuator reading 0–60 dB in steps of 0·5 dB, (iii) a frequency converter and (iv) an i.f. amplifier. The received pulse after detection is displayed

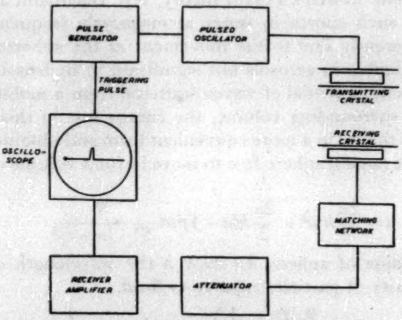


Figure 1. Block diagram of pulse apparatus.

on a double-beam oscilloscope using a time-base triggered from the pulse-generating circuit.

The first term in equation (1) represents a redistribution, on scattering, of the sound energy due to the presence of the sphere and the second term an energy loss due to viscous effects at the surface of the sphere.

At megacycle frequencies with a much less than λ the second term predominates. Urick investigated the importance of this term experimentally and got results in agreement with the predictions of equation (1) but, under the conditions of his experiment, the scattering effect was too small to measure.

The more general case of scattering by particles comparable in radius with the wavelength of the sound does, however, give much greater weight to this scattering term and it was this more general case which was investigated by us. It will appear that the theoretical equations derived reduce to the Lamb–Urick equation if a is made small compared with λ.

When the radius of the sphere approaches the wavelength, calculation of the scattering becomes difficult but Lowan *et al.* (1946) have shown that the attenuation produced by such a rigid sphere on an impinging sound beam is given by

$$\alpha = \frac{\lambda^2}{2\pi} \sum_{m=0}^{m=\infty} (2m+1) \sin^2 F_m\gamma \qquad \ldots\ldots(2)$$

where F_m is a function of $\gamma(=2\pi a/\lambda)$ whose value may be expressed in Bessel functions. They found that for values of γ up to 10 the series in the above summation converged rapidly and they calculated and tabulated the cognate values of $F_m\gamma$.

Taking this expression for α it remained to make allowance for (*a*) the movement of the sphere in the liquid and (*b*) the viscous losses due to the friction at the surface of the sphere. Calculation based on Lamb's treatment showed that movement of the sphere affected only the first-order term in the expansion, which had to be multiplied by Lamb's factor $\{(\sigma-1)/(\sigma+\frac{1}{2})\}^2$. Although Lamb applied the condition $a<\lambda$ it will appear that his 'correction' is still adequate for values of γ up to unity. Similarly the viscous term quoted by Urick appears to be valid for values of γ up to unity.

The receiving quartz crystal is supported by a spring-loaded carriage moving on a screw thread having 20 threads to the inch and a dial enabling readings of one hundredth turn to be made. All readings were taken with the carriage travelling in one direction to eliminate backlash.

Three 'levelling' screws on the backs of the watertight holders containing the crystals enabled their faces to be set accurately parallel. This setting was made by adjusting the screws until maximum acoustic response was attained.

Readings were taken by moving the crystals apart to the maximum required distance, setting the received signal to a reference line on the oscilloscope, increasing the electrical attenuation and then reducing the separation until the signal was brought down to its former level by decrease of the acoustic attenuation. By this means a graph of distance moved against attenuation may be plotted and from the slope of this the absorption coefficient calculated.

Some divergence of the sound beam was encountered particularly at the lower frequencies; an empirical correction was made for this from a series of measurements in distilled water for which the absorption coefficient is accurately known.

For wavelength measurements a slightly different transmitter was used. In this a continuous oscillator fed a gating circuit operated by the multivibrator. The phase difference between the pulse passing through the suspension and that of continuous oscillation fed directly from transmitter to receiver was observed. This phase difference varies with path length, the distance between successive 'in phase' positions being equal to the wavelength. In combination with frequency measurements, using a heterodyne wavemeter, velocity measurements, consistent to 1 in 1000, could be readily obtained. This method of measurement has proved very satisfactory in these experiments, where there is a large amount of mechanical vibration.

2.2. *Tank for Suspensions*

A tank, having a U-shaped cross section, was constructed. It is 14 in. long and 5 in. wide at the top tapering to a curve of 3 in. diameter at the bottom. This is fitted with a four-bladed paddle extending along its whole length and having only a working clearance at the bottom. The axle passes through a watertight bearing in one end of the tank and the paddle can be rotated by an electric motor. With this system particles of up to 200 μ diameter may be kept in suspension with only low speeds of stirring.

Measurements in water alone showed that turbulence caused by the stirrer had no effect on the absorption. Only at speeds high enough to cause air bubbles to be dragged into the water were anomalous effects observed. These speeds were much higher than those used.

Even with this system the upper part of the suspension, where measurements are taken, has not the concentration predicted from the volume of water and the weight of particles added. It is therefore necessary to sample the suspension in the acoustic path. A known volume of the particle + water mixture is sucked out and, after evaporating to dryness, the residue is weighed and the concentration calculated.

A likely source of error was thought to be that due to air trapped on the surface of the spheres. All such air was removed by evacuating the air space above a waterglass mixture, and no change of absorption could be observed. As a precaution all samples of particles were kept under water to ensure thorough wetting.

§ 3. RESULTS IN MONODISPERSE SUSPENSIONS OF GLASS SPHERES

Commercially obtained glass particles have been used to prepare monodisperse systems. They are very nearly spherical and the mean diameter of each sample has been determined by measurement of several hundred individual particles using a comparator.

For any one particle size at any one frequency the absorption due to suspensions of several different concentrations was measured. This was done by adding the particles to the water in several approximately equal quantities and taking a series of measurements after each addition. The quantity of particles added each time was largely a matter of experience, being limited at lower frequencies by the quantity available and at high frequencies by the maximum absorption which the apparatus could conveniently measure.

Measurements on suspensions of particles all having the same size have been performed at six frequencies; 1, 1·5, 2, 3, 6 and 10 Mc/s.

As a specimen of the results, figure 2 shows the attenuations in the 95 μ suspensions in water at 25°c and 1·5 Mc/s. (At this frequency, the practical correction for divergence amounted to 0·8 dB for 10 cm separation of transmitter and receiver. For comparison, the coefficient for pure water is 0·001 at 1 Mc/s and 0·05 at 10 Mc/s.)

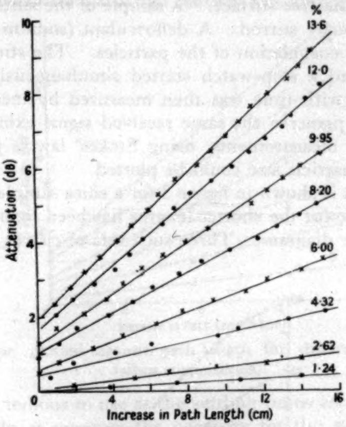


Figure 2. Attenuation in a monodisperse suspension of glass spheres (95 μ) at 1·5 Mc/s and various concentrations.

Table 1 shows the variation of attenuation with concentration and with mean diameter. Results at other frequencies were similar.

Table 1. Measurements at 1·5 Mc/s

Particle diameter (μ)	95	100	110	117	155
Frequency (Mc/s)	1·50	1·50	1·50	1·50	1·50
Wavelength (cm)	0·100	0·100	0·100	0·100	0·100
Attenuation (dB cm^{-1} for 1% vol. conc.)	0·037	0·045	0·054	0·061	0·111
Max. concentration measured (%)	13·6	11·5	10·0	9·1	9·0

Up to a volume concentration of 15 % the increase in velocity of sound was linear, rising in this range from 1498 m sec^{-1} to 1510 m sec^{-1} at a temperature of 25°c.

§ 4. RESULTS IN POLYDISPERSE SUSPENSIONS OF SILICA PARTICLES

Commercially produced glass spheres were not available with diameters below 95 μ and the preparation of large samples of small particles of uniform diameter is very difficult. It was necessary to perform measurements on particles of small diameter to enable a range of frequency to be used.

Urick (1948) in his measurements on viscous absorption used a range of particle size and took a mean value for the diameter. He obtained good agreement with theory. However, in the present case, the absorption due to scattering is expected to vary with radius so the averaging process is very limited in accuracy. Consideration was therefore given to the possibility of utilizing the dependence of the rate of settling of particles in a liquid on the radius to measure the absorption due to a particular size.

For the determination of the particle size distribution in a sample, the optical sedimentation method was used following that described by Richardson (1933, 1943) and Rose (1953).

For the ultrasonic measurements the depth of water in the tank was increased so that the beam between the crystals traversed the suspension at a depth of about 10 cm below the free surface. A sample of the silica was then placed in the tank and vigorously stirred. A deflocculant (sodium hexametaphosphate) was used to prevent coagulation of the particles. The stirring motor was then abruptly stopped and a stop-watch started simultaneously. The decrease in acoustic absorption with time was then measured by increasing the electrical attenuation so as to preserve the same received signal exhibited on the oscilloscope. From these measurements, using Stokes' law, a graph of decrease in attenuation against particle size could be plotted.

One set of curves is shown in figure 3 for a silica sample in water at 10 Mc/s. The attenuation scale for the shorter lengths has been reduced in order to show all the curves on one diagram. Three such sets of curves were plotted for any one sample.

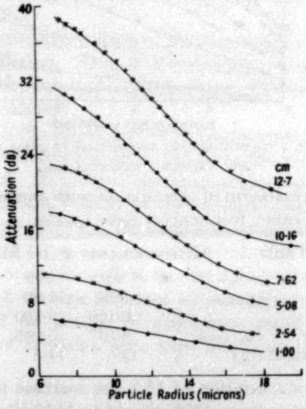


Figure 3. Attenuation in polydisperse suspensions of silica plotted against particle size for different path lengths.

In assessing the average values from these, an arbitrary zero was used to eliminate any errors due to the incorrect adjustment of the pulse height for the first measurement. In the case illustrated the 8μ reading was taken as zero. By measuring the average increase in attenuation between 8 and 9μ for the three sets of curves at various path lengths, a graph of attenuation against path length could be plotted. Similarly the increase in attenuation between 8 and 10μ, 8 and 11μ, etc., could be measured and further graphs plotted. These are shown in figure 4 and as expected from the work on monodisperse systems the graphs are straight lines. The attenuation for the particles between say 12 and 13μ may then be calculated from the difference in slope between the two graphs plotted for those sizes.

The concentration of all the particles in the suspension was then measured by the sampling method and the particle size distribution by the photo-extinction method. From these measurements the volume concentration of particles of any size may be calculated. Knowing this concentration and the absorption in dB cm^{-1} from graphs of the type given in figure 4, the absorption for a 1% volume concentration of the particular particle size may be calculated.

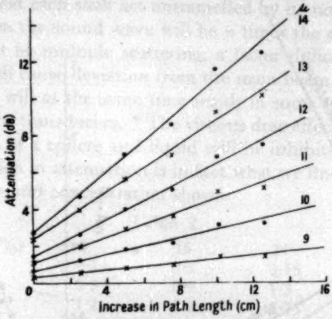


Figure 4. Attenuation plotted against path length for different (mean) particle sizes (silica suspensions).

The method is tedious to use as the settling times are long but it has certain advantages. Firstly, it removes the necessity for the grading of the particles according to size. Secondly, it enables the absorption due to many particle sizes to be measured in one experiment, and thirdly, the pulse is very steady and its height may be accurately adjusted to the reference level.

The measurements were made at 6 and 10 Mc/s.

Discussion.

The fact that at the low concentration so far used the ultrasonic attenuation is directly proportional to concentration shows that interaction between the particles may be neglected (as assumed in the theory).

From the theoretical formula (3) it may be seen that, apart from the last term due to viscous loss, the absorption for a given number of particles per cm^3 rises with γ, so for ease in presentation of the experimental results the corresponding viscous loss, which exhibits a complex dependence upon frequency and particle radius, has been deducted from each reading. This process was assumed to be legitimate in view of the work of Urick who confirmed experimentally the theoretical values of the viscous loss, plotted together with the scattering term in figure 5. When γ is comparable with unity, the 'correction' for viscous loss can be safely neglected.

Results are plotted as $\alpha\lambda$ against γ for monodisperse (figure 6) and polydisperse systems (figure 7), the full curve in each being given by the formula (3). The agreement is within experimental accuracy.

§ 5. RESULTS IN SEDIMENTS, ARTIFICIAL AND NATURAL

Measurements were made in the apparatus with concentrations of glass spheres up to 40%, at which maintenance of the suspension by stirring became difficult. In addition attenuation was measured at the maximum packing fraction of 70%, to which the term 'suspension' can hardly be applied. Rather was the system then a loosely compacted sediment in the water. For these the ultrasonic beam was transmitted vertically through the water with the glass spheres lying in a layer to a measured height above the transmitter, the receiver

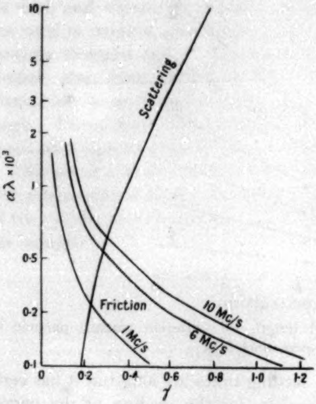


Figure 5. Calculated viscous and scattering losses for various values of $\gamma(=2\pi a/\lambda)$ at several frequencies ($C=1\%$; $\sigma=2{\cdot}6$).

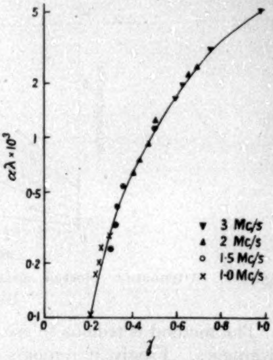


Figure 6. Attenuation per wavelength plotted against γ in monodisperse suspensions—full line, theory.

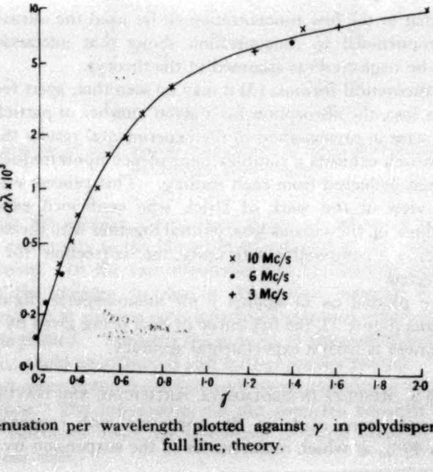


Figure 7. Attenuation per wavelength plotted against γ in polydisperse suspensions—full line, theory.

lying in the water above. Pulses were then transmitted between the fixed transducers while the depth of layer of sediment was increased by steps. The attenuation–depth plots were as before straight lines, from the slopes of which the attenuation per centimetre was calculated.

Of course, with these large concentrations, the conditions of Sewell's theory are not fulfilled. There will be 'interaction' between the particles, whereas the theory requires that each shall act untramelled by its neighbours, so that the effect of *n* particles on the sound wave will be *n* times the effect of one.

There will in fact be multiple scattering, a factor difficult to introduce into any theory, which will cause deviation from the main beam to an extent difficult to allow for, since it will at the same time result in some of the multi-scattered energy returning to the transducers. The viscous drag effect will also be lessened since relative motion of a sphere and liquid will be inhibited by its neighbours. The expected reduction in attenuation is in fact what we find, as an inspection of table 2 of attenuation and concentration shows.

Table 2

Volume concentration (%)		8	16	24	32	40
Attenuation (dB cm^{-1}) at	1 Mc/s	0·3	0·5	0·75	1·0	1·2
	1·5 Mc/s	0·5	0·9	1·3	1·7	1·9
	2 Mc/s	1·2	2·2	2·75	2·9	2·95

Some measurements of attenuation were also made in natural sediments of sifted sand from the seashore, since the attenuation of ultrasonics in such materials is a matter of importance in the practice of underwater sound.

Figure 8 shows results at various frequencies in such natural sediments graded in three size ranges, A, B, C, with mean diameters 450, 180, 120μ. For comparison, some results in an artificial sediment of glass spheres which had almost the same mean as specimen C are marked (as crosses) on the graph. The concurrence of the artificial and the natural sediment is rather interesting, firstly because sand grains are not, of course, spherical, and secondly since it is well known that the material in a natural sediment posseses visco-elastic properties—often, thixotropy, notably in quicksands—which the system of glass spheres did not show.

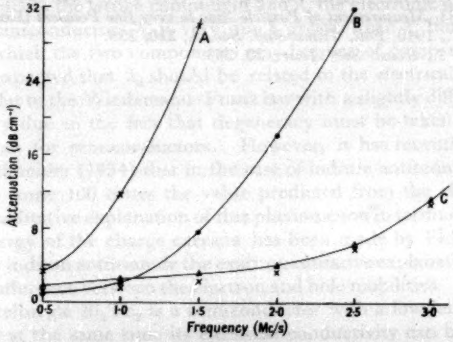


Figure 8. Attenuation plotted against frequency in natural sediments A, B, C. (The crosses beside curve C refer to an artificial sediment.)

One further point is to be remarked on the graphs of figure 8, which show also the variation of attenuation with frequency. At high frequencies the Rayleigh scattering law (α proportional to f^4) seems to be obeyed by these sediments but at low frequencies the 'absorption' seems to be directly proportional to frequency.

It appears that an acoustic method might be used for particle size analysis in conjunction with a sedimentation cell, without more experimental difficulty than is involved in an optical measurement. The main experimental advantage of this method of analysis appears to lie in the pulsing operation, whereby mechanical radiation forces on the particles and resulting convection are avoided since the source is not continuously operated. The apparatus would, however, be somewhat larger than that used for the optical method, requiring a larger sample of powder for analysis.

ACKNOWLEDGMENTS

The authors are pleased to acknowledge the help and advice which Dr. L. E. Lawley contributed to the design and development of the apparatus. They are also grateful to the Royal Naval Scientific Service for a maintenance grant to one of them (J.B.) while this work was being carried out.

REFERENCES

HARTMANN, C. R., and FOCKE, A. B., 1940, *Phys. Rev.*, **57,** 221.
LAIDLER, T. J., and RICHARDSON, E. G., 1938, *J. Acoust. Soc. Amer.*, **19,** 217.
LAMB, H., 1916, *Hydrodynamics* (4th Edn), 507.
LOWAN, A. N., (Ed.), 1946, *Scattering of Radiation from circular Cylinders and Spheres.* (Rep. U.S. Navy. Office of Research and Investigations.)
RAYLEIGH, Lord, 1878, *Theory of Sound*, **II,** 272.
RICHARDSON, E. G., 1933, *J. Agric. Sci.*, **24,** 457; 1943, *Proc. Phys. Soc.*, **55,** 48.
ROSE, H. E., 1953, *Measurement of Particle Size in very fine Powders* (London: Constable).
SEWELL, C. J. T., 1910, *Phil. Trans. Roy. Soc.* A, **210,** 239.
URICK, J., 1948, *J. Acoust. Soc. Amer.*, **20,** 283.

The Thermal Conductivity of Bismuth Telluride

By H. J. GOLDSMID

The Research Laboratories of The General Electric Company, Ltd., Wembley, Middx

MS. received 16th September 1955

Abstract. Measurements of the thermal conductivity of bismuth telluride have been made in the range of temperature 150°K to 300°K. It has been found that the electronic contribution to the conduction of heat is considerably greater for specimens in which the charge carriers are intrinsically excited than for specimens in which most of the carriers arise from an impurity concentration; this can be explained by the recently proposed theory which takes into account the transfer of ionization energy down a temperature gradient.

---

§ 1. Introduction

Heat may be conducted through a solid either by elastic waves in the crystal lattice or by free charge carriers. In metals it is generally sufficient to consider only the contribution to the flow of heat from the free electrons; hence the Wiedemann–Franz law relating electrical and thermal conductivities. On the other hand in semiconductors it is also necessary to take into account the contribution from the elastic waves, and the total thermal conductivity may be expressed as

$$\lambda = \lambda_0 + \lambda_e \qquad \ldots\ldots(1)$$

where λ_0 represents the lattice component and λ_e the electronic component.

In most semiconductors λ_0 is very much greater than λ_e but there are a few materials in which the two components may become of comparable magnitude. It might be expected that λ_e should be related to the electrical conductivity by a relation similar to the Wiedemann–Franz law with a slightly different constant of proportionality due to the fact that degeneracy must be taken into account for metals, but not for semiconductors. However, it has recently been noted by Busch and Schneider (1954) that in the case of indium antimonide the electronic component is some 100 times the value predicted from the electron theory of metals. A qualitative explanation of this phenomenon in terms of the transported ionization energy of the charge carriers has been made by Fröhlich and Kittel (1954), but for indium antimonide the exact quantitative explanation is complicated by the large difference between the electron and hole mobilities.

Bismuth telluride Bi_2Te_3 is a semiconductor with a low lattice thermal conductivity and at the same time its electrical conductivity can be quite high. It therefore seemed a suitable material upon which to investigate the contribution to the conduction of heat from the charge carriers.

§ 2. Experimental Procedure and Results

The apparatus used for the absolute measurement of the thermal conductivity is shown in figure 1. The source of heat consisted of an insulated nichrome coil embedded in a copper block, the power supplied being calculated from current and voltage measurements. The specimens were of square cross section and were ground flat and parallel, electroplated, and soldered into position to ensure good thermal contact. The sink consisted of a copper cylinder within which slid a copper block which was soldered to the bottom of the specimen. The temperatures of the heat source and sink were measured by means of copper–constantan thermocouples. All the electrical leads, which were made from 40 s.w.g. wire, passed through a clamp at the same temperature as the sink, and the enclosure was evacuated. The loss of heat from the source, other than through the specimen, was thus small and definite; its value, measured over the full range of temperature without a specimen in place, was very close to the figure predicted from the dimensions of the apparatus, and consequently a reliable correction could be applied.

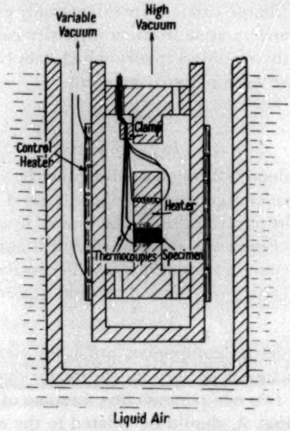


Figure 1. Thermal conductivity apparatus.

Control of the temperature of the sink was obtained by varying the degree of vacuum in the space which separated it from an outer cylinder immersed in liquid air, and by adjusting the current flowing in a nichrome heating coil.

Six samples of bismuth telluride were employed. Specimens 1, 2 and 3 were cut from zone-refined bars with their faces perpendicular to the direction of zoning; the cleavage planes then lay along the direction of measurement. Specimens 4, 5 and 6 were cut from a single crystal; in specimen 4 the measurement was made parallel to, and in specimens 5 and 6 perpendicular to the cleavage planes. The thermoelectric power and electrical conductivity parallel to the cleavage planes were measured at room temperature for samples adjacent to the heat conduction specimens; in the case of specimens 1 and 2 the measurement of electrical conductivity was extended to cover the range of temperature 150°K to 300°K as shown in figure 2. The electrical properties at room temperature are summarized in table 1.

The results from the measurement of thermal conductivity for all the samples are given in figure 3. The importance of the anisotropic effect is illustrated by the low thermal conductivity of specimens 5 and 6 as compared with specimen 4; this

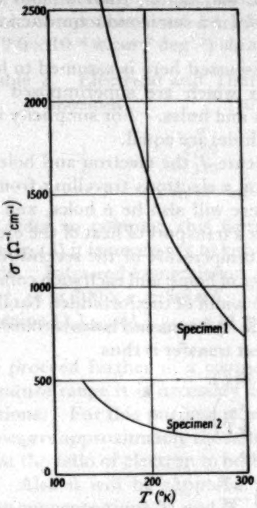


Figure 2. Electrical conductivity plotted against temperature.

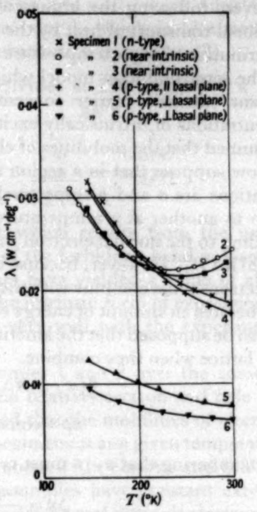


Figure 3. Thermal conductivity plotted against temperature.

effect will not, however, be discussed further. Much more interesting is a comparison of the thermal conductivity of the relatively pure specimens 2 and 3 with the impure specimens 1 and 4. It is remarkable that, whereas at low temperatures the former have an inferior heat conductivity, the opposite is true at room temperature, even though the impure samples still have an appreciably greater electrical conductivity.

Table 1. Electrical Properties at Room Temperature

(1)	(2)	(3)	(4)	(5)	(6)
1	−215	n	750	Impurity	
2	− 65	n	200	Near intrinsic	Adjacent specimens
3	−200	n			
4	+200	p	500	Impurity	Adjacent specimens
5					
6					

(1) Heat conduction specimen; (2) thermo-electric power (μv deg⁻¹); (3) type of conduction; (4) electrical conductivity of adjacent samples parallel to cleavage planes ($\Omega^{-1}\,\mathrm{cm}^{-1}$); (5) mechanism of carrier excitation; (6) remarks.

§ 3. THE ELECTRONIC COMPONENT OF THE THERMAL CONDUCTIVITY

According to the electron theory of metals (Sommerfeld 1928) the electronic component of the thermal conductivity λ_e and the electrical conductivity σ should be related by the expression

$$\frac{\lambda_e}{\sigma} = A\left(\frac{k}{e}\right)^2 T \qquad \ldots\ldots(2)$$

where $A=2$ for non-degenerate conductors and 3·3 for degenerate conductors. However, following the argument of Fröhlich and Kittel, there should be an additional transport of heat by the charge carriers in a semiconductor due to their redistribution along a temperature gradient.

The semiconductor model which will be discussed here is assumed to have a constant extrinsic carrier concentration upon which are superimposed equal concentrations of intrinsically excited electrons and holes. For simplicity it will be assumed that the mobilities of electrons and holes are equal.

Now suppose that in a region at a temperature T the electron and hole concentrations are n and p respectively. Then for n electrons travelling from this region to another at a temperature $T-\Delta T$ there will also be p holes, and thus, according to the normal electron theory, there is a transport of heat of the order of $\frac{3}{2}(n+p)\boldsymbol{k}\Delta T$. However, because of the lower temperature of the second region, Δn electrons must combine with the same number of holes, and each such combination liberates an amount of energy equal to ϵ_g the width of the forbidden band. It will also be supposed that the kinetic energy of the electrons and holes is transferred to the lattice when they combine. The total heat transfer is thus

$$\tfrac{3}{2}(n+p)\boldsymbol{k}\Delta T+\Delta n(\epsilon_g+3\boldsymbol{k}T).$$

Now

$$np=\text{const.}\ T^3\exp\left(-\frac{\epsilon_g}{\boldsymbol{k}T}\right)$$

so, remembering that $n-p$ must remain constant,

$$\Delta n=\frac{np}{n+p}\left(\frac{\epsilon_g}{\boldsymbol{k}T}+3\right)\frac{\Delta T}{T}.$$

Thus the ratio of the total electronic component of the thermal conductivity $(\lambda_e)_{\text{total}}$ to the value $(\lambda_e)_{\text{normal}}$, which is given if the transport of ionization energy is ignored, is

$$\frac{(\lambda_e)_{\text{total}}}{(\lambda_e)_{\text{normal}}}=1+\frac{2}{3}\frac{np}{(n+p)^2}\left(\frac{\epsilon_g}{\boldsymbol{k}T}+3\right)^2. \qquad \ldots\ldots(3)$$

In the intrinsic case where $n=p$ this simplifies to

$$\left[\frac{(\lambda_e)_{\text{total}}}{(\lambda_e)_{\text{normal}}}\right]_{\text{intrinsic}}=1+\frac{1}{6}\left(\frac{\epsilon_g}{\boldsymbol{k}T}+3\right)^2.$$

§ 4. DISCUSSION OF RESULTS

The apparent anomaly in the thermal conductivity of bismuth telluride may now be explained qualitatively. At the lowest temperatures the normal electronic component predominates for all specimens, so those with the higher electrical conductivity, numbers 1 and 4, also have the higher thermal conductivity. At room temperature, however, the intrinsic carrier concentration is greater for specimens 2 and 3, and the consequent transport of ionization energy is sufficient to raise the thermal conductivity of these specimens above that for samples 1 and 4.

This explanation is confirmed quantitatively by comparing the thermal conductivity of sample 1 at 150°K with that of sample 2 at 300°K. The relevant data are listed in table 2. The total thermal conductivity of specimen 1 at 150°K is $3{\cdot}16\times10^{-2}\,\text{w cm}^{-1}\,\text{deg}^{-1}$ and it is assumed that the electronic component of the thermal conductivity may be obtained from equation (2), so that its value is $6{\cdot}0\times10^{-3}\,\text{w cm}^{-1}\,\text{deg}^{-1}$ if A is taken to be equal to 2, the non-degenerate figure. The lattice component at 150°K is thus $2{\cdot}56\times10^{-2}\,\text{w cm}^{-1}\,\text{deg}^{-1}$. If, as is usual, it is assumed that the lattice component varies inversely with the absolute temperature, its value at 300°K is $1{\cdot}28 \times 10^{-2}$ w cm^{-1} deg^{-1}. Since the total thermal conductivity of specimen 2 at 300°K is $2{\cdot}44 \times 10^{-2}$ w cm^{-1} deg^{-1} the electronic component must be $1{\cdot}16 \times 10^{-2}$ w cm^{-1} deg^{-1}, which is 13 times the value ($9{\cdot}0 \times 10^{-4}$ w cm^{-1} deg^{-1}) obtained from substitution in relation (2).

Table 2. Electrical and Thermal Conductivities at 150°K and 300°K

Specimen	150°K		300°K	
	σ	λ	σ	λ
1	2700	0·0316	730	0·0187
2	370	0·0278	200	0·0244

σ in Ω^{-1} cm^{-1}; λ in w cm^{-1} deg^{-1}.

In order to compare this factor with that which results from the use of expression (3) it is necessary to know the width of the forbidden band in bismuth telluride. Infra-red transmission measurements have yielded a value of about 0·15 ev†, so at 300°K $\epsilon_g \simeq 5{\cdot}8\,\boldsymbol{k}T$. Thus using the intrinsic form of expression (3) one obtains $(\lambda_e)_{\text{total}}/(\lambda_e)_{\text{normal}} = 14$ which agrees very well with the experimental value.

To proceed further in a comparison of samples 1 and 2 over the complete temperature range it is necessary to estimate the relative electron and hole concentrations. For this purpose it will be assumed that the mobilities of electrons and holes are approximately the same for both specimens at any given temperature, and that the ratio of electron to hole mobility is about 2 : 1 (Vlasova and Stil'bans 1955). Also it will be supposed that the two samples have constant extrinsic electron concentrations N_1 and N_2, to which are added equal intrinsic electron and hole concentrations n_1 and n_2 respectively. Then at a given temperature

$$\frac{\sigma_1}{\sigma_2} = \frac{(N_1+n_1)+\frac{1}{2}n_1}{(N_2+n_2)+\frac{1}{2}n_2}$$

and if non-degeneracy is assumed, which can only be approximately true in view of the comparatively low thermoelectric power of sample 1, $(N_1+n_1)n_1 \simeq (N_2+n_2)n_2$.

Substitution in these expressions at 150°K and 300°K yields values of N_1 and N_2 equal to $6{\cdot}6\,n_0$ and $0{\cdot}9\,n_0$ respectively, n_0 being the intrinsic electron or hole concentration in bismuth telluride at 300°K. Using these values table 3 has been drawn up.

Table 3. Estimation of Electronic Component of the Thermal Conductivity

Specimen 1

T(°K)	$\dfrac{\epsilon_g}{\boldsymbol{k}T}$	$\dfrac{n_1(N_1+n_1)}{(N_1+2n_1)^2}$	$\dfrac{\lambda_{\text{total}}}{\lambda_{\text{normal}}}$	σ_1 (Ω^{-1} cm^{-1})	λ_{e1} (w cm^{-1} deg^{-1})
150	11·6	$8{\cdot}7 \times 10^{-6}$	1·001	2700	0·0060
200	8·7	$3{\cdot}7 \times 10^{-4}$	1·034	1600	0·0049
250	7·0	$3{\cdot}9 \times 10^{-3}$	1·26	1040	0·0049
300	5·8	$2{\cdot}1 \times 10^{-2}$	2·09	730	0·0068

Specimen 2

T(°K)	$\dfrac{\epsilon_g}{\boldsymbol{k}T}$	$\dfrac{n_2(N_2+n_2)}{(N_2+2n_2)^2}$	$\dfrac{\lambda_{\text{total}}}{\lambda_{\text{normal}}}$	σ_2 (Ω^{-1} cm^{-1})	λ_{e2} (w cm^{-1} deg^{-1})
150	11·6	$4{\cdot}7 \times 10^{-4}$	1·067	370	0·0009
200	8·7	$1{\cdot}9 \times 10^{-2}$	2·73	245	0·0020
250	7·0	$1{\cdot}15 \times 10^{-1}$	8·7	200	0·0065
300	5·8	$2{\cdot}08 \times 10^{-1}$	11·8	200	0·0106

---

† Private communication from I. G. Austin of these Laboratories.

From table 3 it may be seen that the calculations lead one to suppose that the electronic component of the thermal conductivity should be the same for both specimens at a temperature between 200°K and 250°K. Reference to figure 3 shows that this is the case at 225°K.

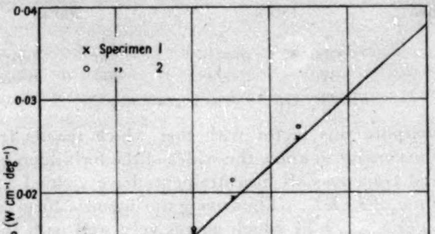


Figure 4. Lattice component of the thermal conductivity.

By subtracting the calculated electronic component from the total measured thermal conductivity the lattice component should be obtained. This procedure has been used in obtaining figure 4. It may be seen that the values of λ_0 obtained in this way for the two samples lie reasonably close to each other and follow the expected inverse temperature variation.

§ 5. CONCLUSIONS

The close agreement between the predicted values for the lattice component of the thermal conductivity, for the two specimens considered, confirms that the energy of excitation of the charge carriers should be taken into account in calculating the electronic component of the thermal conductivity of semiconductors and, in particular, it appears that this is the reason for the anomalously high thermal conductivity of pure bismuth telluride. A more complete theory of the extra electronic component should take into account the difference between the mobilities of electrons and holes, and the importance of degeneracy in the calculations should also be considered, but the effect of these refinements on the simple theory which has been presented here must be quite small, in view of the good correlation which has been achieved.

The measurements show that the lattice component of the thermal conductivity of bismuth telluride, parallel to the basal plane, varies inversely with absolute temperature over the experimental range, obeying the relation

$$\lambda_0 = 3{\cdot}8/T \text{ w cm}^{-1}\text{ deg}^{-1}.$$

ACKNOWLEDGMENTS

The author is indebted to Dr. R. Wolfe and Mr. R. D. Groves for making the measurements of electrical conductivity.

REFERENCES

BUSCH, G., and SCHNEIDER, M., 1954, *Physica*, **20**, 1084.
FRÖHLICH, H., and KITTEL, C., 1954, *Physica*, **20**, 1086.
SOMMERFELD, A., 1928, *Z. Phys.*, **47**, 1.
VLASOVA, R. M., and STIL'BANS, L. S., 1955, *Z. Tekh. Fiz. SSSR*, **25**, 569.

Magnetic Properties of some Anhydrous Chlorides of the Iron Group

I: Magnetic Resonance

By J. W. LEECH and A. J. MANUEL†
Department of Physics, Queen Mary College, London

MS. received 1st *September* 1955

Abstract. Measurements of paramagnetic resonance absorption in $CrCl_3$ and $NiCl_2$ in the temperature range 13·15°K to 293°K are reported. The height of the absorption curve decreases and the line width increases as the temperature is lowered through the region of the cryomagnetic anomalies in these salts. The results are discussed in part II, where consideration of other experimental work suggests that the magnetic transition in these salts at low temperatures is essentially ferromagnetic rather than antiferromagnetic.

$FeCl_2$ and $CoCl_2$ were also examined in the same temperature range but paramagnetic resonance was not observed.

§ 1. INTRODUCTION

THE anhydrous chlorides of the iron group $CoCl_2$, $FeCl_2$, $NiCl_2$ and $CrCl_3$ show similar cryomagnetic anomalies. They all have a specific heat anomaly at low temperatures associated with changes in their magnetic properties. Above the temperature of the specific heat anomaly (henceforth referred to as the transition temperature T_λ) they follow to a first approximation a Curie–Weiss law with a positive value of θ if this law is written $\chi = C/(T-\theta)$. This corresponds to a positive molecular field on the Weiss theory and should lead to ferromagnetism below T_λ. Below this temperature however, they differ from normal ferromagnetic substances in showing exceptional magnetic hardness and almost negligible remanences, with susceptibilities of the same order as paramagnetic substances. The susceptibilities for low fields show a maximum at T_λ for $CoCl_2$, $FeCl_2$, and $NiCl_2$ and this has led them to be classed, in recent reviews (Lidiard 1954, Nagamiya, Yosida and Kubo 1955) as antiferromagnetic. $CrCl_3$ has no maximum in the susceptibility and has not been looked on as antiferromagnetic, but is of similar crystal structure.

A summary of the relevant data for these salts is given in table 1.

An investigation of the variation of magnetic resonance absorption in these salts, as they are cooled through their transition temperatures, was undertaken to compare the behaviour of $CrCl_3$ with the others.

Paramagnetic resonance absorption occurs by the absorption of a quantum of microwave radiation in stimulating transitions between the Zeeman levels in an atom or ion. The condition for this to occur is that

$$h\nu = g\beta H \qquad \ldots\ldots(1)$$

where ν is the frequency of the incident radiation, g the spectroscopic splitting factor, β the Bohr magneton and H the applied magnetic field causing splitting of the spin levels. In a ferromagnetic or antiferromagnetic substance the effective field acting on the ion is no longer H but depends on the internal field. This causes a change in the value of the applied field at which resonance takes place. In a ferromagnetic substance resonance is observed at lower applied fields than is given by the simple formula (1) above, due to the internal field. The situation is more complicated in antiferromagnetic substances, but in cases where the transition temperature is more than a few degrees above absolute zero the antiferromagnetic internal field shifts the magnetic resonance line beyond the range of present day experimental apparatus, and all that is observed is a disappearance of the paramagnetic resonance line as the substance is cooled through the transition temperature.

---

† Now at Bedford College, London

Table 1

	P_{eff}	θ(°K)	T_λ (°K)	$T(\chi_{\mathrm{m}})$(°K)
$FeCl_2$	5·38 (d)	20·4 (b)	23·5 (e)	24·0 (d)
		48·0 (d)		
$NiCl_2$	3·32 (d)	67·0 (a)	49·6 (f)	50·0 (a)
		68·2 (d)	52·35 (g)	
$CoCl_2$	5·29 (d)	20·0 (a)	24·9 (f)	25·4 (c)
		19·0 (c)		
$CrCl_3$	3·90 (d)	32·5 (a)	16·8 (f)	—
		27·0 (d)		

P_{eff} = effective Bohr magneton number; θ = paramagnetic Curie temperature from Curie–Weiss law; T_λ = temperature of specific heat anomaly; $T(\chi_{\mathrm{m}})$ = temperature of maximum susceptibility.

(a) Woltjer (1925); (b) Woltjer and Wiersma (1929); (c) Shalyt (1938); (d) Bitter, Starr and Kaufmann (1940); (e) Trapeznikowa and Schubnikow (1935); (f) Trapeznikowa, Schubnikow and Miljutin (1936); (g) Busey and Giauque (1952).

Polycrystalline specimens were examined in the temperature range 13·15°K to 293°K. The results of these measurements are given in detail in § 3 and are discussed in part II, together with other experimental evidence which suggests that the magnetic transition in these salts is to a ferromagnetic state at low temperatures.

§ 2. EXPERIMENTAL APPARATUS AND METHOD

2.1. *General Description of Apparatus*

The apparatus can best be described by considering the three principal parts, the microwave system, the hydrogen liquefier and the magnet.

A block diagram of the microwave circuit is shown in figure 1. Power from the klystron oscillator O fed from stabilized d.c. supplies was coupled into the primary waveguide in which was located the attenuator A_1 and coupled out at the transition T_1 to a coaxial line to the cavity resonator R. A second coaxial line carried the transmitted power from the resonator and coupled it into the detector waveguide at T_2. The detector waveguide contained the attenuator A_2 and the crystal detector X_2. The current in X_2 was indicated by the galvanometer G_2. G_2 was a sensitive suspended coil galvanometer, used with a curved scale at a distance of 2 metres. The galvanometer and scale combination was calibrated and found to be linear over the range of the scale. The galvanometer and its connection to the crystal X_2 were carefully screened. A secondary waveguide was coupled by a 30 dB Bethe hole coupler at D. The absorption frequency meter F was mounted on this secondary waveguide at the other end of which was mounted the crystal detector X_1. The galvanometer G_1 indicated the current in X_1 and was used to monitor the power level in the primary waveguide.

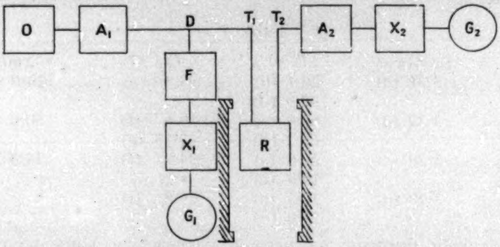


Figure 1. **Block diagram of experimental apparatus.**

Both transitions and the crystal mounts were matched with non-contact short-circuit plungers, and matching screws were provided at the crystal detector mounts.

The resonator of circular cross section was designed to support the H_{11} and suppress the E_{01} mode over the tuning range of the klystron (9550–9200 Mc/s), by choosing the diameter to be 2·31 cm. This gives a free space cut-off wavelength λ_c for the H_{11} mode and E_{01} modes of 3·94 cm and 3·08 cm respectively. For a cylindrical resonator in the H_{11} mode the electric field vector is a minimum at the end plate, while the magnetic field vector is a maximum and directed along a diameter defined by the position of the coupling loops. With a specimen mounted at the centre of the end plate, the most favourable conditions for paramagnetic resonance absorption are obtained when the oscillatory magnetic vector is arranged to be perpendicular to the static magnetic field. The end of the resonator distant from the coupling loops was closed by a non-contact choke tuning piston designed for a free space wavelength of 3·2 cm, and mounted on a fine screw thread in the barrel of the resonator. Calculation of the unloaded *Q* for such a design gave a value for a brass cavity of 10 000. The resonator was surrounded by and integral with a vacuum-tight cooling jacket whose overall length was over three times that of the resonator space, the coaxial leads and thermocouples attached to the resonator being shielded by the cooling jacket. The cooling jacket was also the liquefying chamber for the hydrogen liquefier.

The miniature hydrogen liquefier was similar in construction to the helium liquefier described by Chester and Jones (1953) without the helium stage. The lower part of the vacuum casing was of reduced cross section and contained the resonator and its combined cooling jacket. The liquefier was surrounded by a long-tailed Dewar vessel containing liquid oxygen and the magnet was rolled into a previously marked position so that the specimen in the resonator was on the axis of the magnet pole pieces. Temperatures in the range from 20·4°K to 90°K could be obtained by operating the liquefier inefficiently by reducing the pressure of the incoming hydrogen, and by partially bypassing the cooled hydrogen from the resonator cooling jacket to reduce the efficiency of the heat exchanger. Temperatures below 20·4°K were obtained by boiling hydrogen under reduced pressure in the cooling jacket. Temperatures above 20·4°K were measured with a copper–constantan thermocouple attached to the resonator opposite the specimen, within the cooling jacket. The e.m.f. of the thermocouple was determined with its warm junction in melting ice and the cold junction in liquid oxygen, liquid nitrogen, and liquid hydrogen in turn, and intermediate temperatures obtained from a table calculated from the equation $E = At + Bt^2 + Ct^3$ using the three calibration points to determine A, B and C. Temperatures below 20·4°K were derived from the vapour pressure of the hydrogen boiling in the cooling jacket, using the data of Linder (1950).

The electromagnet provided a field up to 5500 oersteds in a gap of 6 cm. The pole pieces were shimmed to render the field more homogeneous (Rose 1938). The current was supplied from the d.c. mains supply, stabilized by a battery of accumulators in the circuit described by Potter (1934). The magnet current was measured with a standard resistance and potentiometer circuit. The magnet field was calibrated with a search coil and ballistic galvanometer and standardized on the paramagnetic resonance line of $\alpha\alpha$diphenyl β picryl hydrazl, a specimen of which was mounted in the same position in the resonant cavity as the salts under investigation. The accuracy of the field calibration was estimated at 0·5% with an inhomogeneity over the specimen of 0·1%.

2.2. *Types of Measurement and Procedure*

Three types of measurement were made: (*a*) The variation in power transmitted by the cavity containing the specimen in the absence of an external static magnetic field was observed as the temperature was allowed to rise slowly. A similar measurement was also made on the empty cavity. (*b*) The variation in power transmitted by the cavity as the static field was increased from zero to 5500 oersteds was measured at a series of steady temperatures. (*c*) For $NiCl_2$ the variation with increasing temperature of the power transmitted by the cavity containing the specimen, in a magnetic field to give maximum absorption, was also measured.

From (*a*) the variation of the relative absorption in zero field with changing temperature could be calculated; from (*b*) the variation of relative absorption with field at a given temperature, i.e. the magnetic resonance absorption spectrum, was obtained; (*c*) gave for $NiCl_2$ a continuous curve of the absorption maximum against temperature for the range from liquid hydrogen to liquid oxygen temperatures.

In all measurements care was taken to ensure that the klystron was tuned to the resonant frequency of the cavity for each observation and that the power input to the primary waveguide was constant for a series of measurements. This was achieved by using the attenuator A_1 and monitor galvanometer G_1 of figure 1. The temperature was kept constant and checked at each observation in a series of measurements of the type (*b*). Measurements were made at different power levels in the detector crystal to verify that the departures of the crystal from square law rectification were within other experimental errors.

2.3. *Treatment of Observations*

It can be shown that if, for a given input power, P_0 is the power transmitted by a resonant cavity at its resonant frequency in the absence of magnetic absorption and P_A that transmitted at resonance when in addition magnetic absorption takes place then

$$\chi'' \propto \left(\frac{P_0}{P_A}\right)^{1/2} - 1$$

where χ'' is the imaginary part of the complex susceptibility. If the power is measured as in these experiments by the current in a crystal detector operating in the square law region then

$$\chi'' \propto \left(\frac{d_0}{d_A}\right)^{1/2} - 1$$

where d_0 is the galvanometer deflection in the absence of absorption and d_A that in the presence of magnetic absorption. The graph of the function $(d_0/d_A)^{1/2} - 1$ plotted against applied magnetic field gives the paramagnetic absorption spectrum for a given frequency ν. Under suitable experimental conditions peaks should be obtained in this curve corresponding to magnetic resonances, the position of the peaks being given by the condition that the quantum of microwave energy $h\nu$ should be equal to the separation of the Zeeman levels in the atom: $h\nu = g\beta H$ where g is the spectroscopic splitting factor and β the Bohr magneton.

It is not always clear what value should be taken for P_0 and it is customary to assume zero absorption in high magnetic fields, i.e. $\chi'' = 0$. Preliminary examination of the results showed that above the temperature of the cryomagnetic anomaly narrow resonance lines were obtained but below this temperature the lines were broad with high zero field absorption. With these broad lines it was doubtful whether at $H = 5500$ oersteds χ'' was zero.

For the narrow lines at temperatures above T_λ, χ'' was assumed zero in zero field. This enabled a curve of variation of relative absorption in zero field against temperature to be constructed. From this curve and the ratio of the peak relative absorptions at 90°K for different specimens, the value of relative absorption in zero field at any temperature for any specimen could be obtained. Hence a galvanometer deflection G_2 corresponding to zero absorption for any set of measurements could be calculated. Measurements made on the empty cavity were used to correct for change in Q with temperature.

2.4. *Preparation of Specimens*

A sealed ampoule of anhydrous chromic chloride was kindly supplied by Dr. P. D. F. Shaw of the Clarendon Laboratory, Oxford. The remaining salts were prepared from the hexahydrates of Analar standard by dehydration under vacuum and vacuum sublimation. The procedures adopted for each salt were the same apart from the temperatures at which sublimation was carried out. This was 800°C for $NiCl_2$, 900°C for $CoCl_2$ and 800°C for $FeCl_2$. A fair quantity of the hexahydrate was dehydrated at a temperature between 100°C and 200°C. The dehydration caused the salts to coagulate into lumps and it was necessary to grind them to a fine powder in a mortar at frequent intervals. When a salt showed no further change in colour during several hours under vacuum at 200°C, the anhydrous powder was transferred to a clear silica sublimation tube which was evacuated. The anhydrous salt was placed at one end of the sublimation tube which was then inserted into the furnace. After a few hours an adequate amount of the sublimed chloride was formed in the cooler parts of the tube. The sublimation tube was then removed from the furnace, allowed to cool and opened. The purified sublimate was gently scraped out and pressed into a small cylindrical pill 4 mm in diameter in a clean brass mould. The thickness of the pill varied from a fraction of a millimetre to about 4 mm for different specimens. The pill was mounted in the resonator with dilute 'Durofix' cement and the resonator sealed in the liquefier casing and evacuated. The time between the opening of the sublimation tube and the evacuation of the liquefier casing was made short as possible (a few minutes) since the anhydrous salts are hygroscopic.

§ 3. RESULTS

Investigations were carried out on the anhydrous chlorides, $CoCl_2$, $FeCl_2$, $NiCl_2$ and $CrCl_3$. Resonance absorption was not observed in $FeCl_2$ or $CoCl_2$ over the range of temperatures from 13·15°K to room temperature, although for $CoCl_2$ a very slight decrease in absorption with increasing field was observed, just above the temperature of the cryomagnetic anomaly. This was not sufficiently large to present significant results.

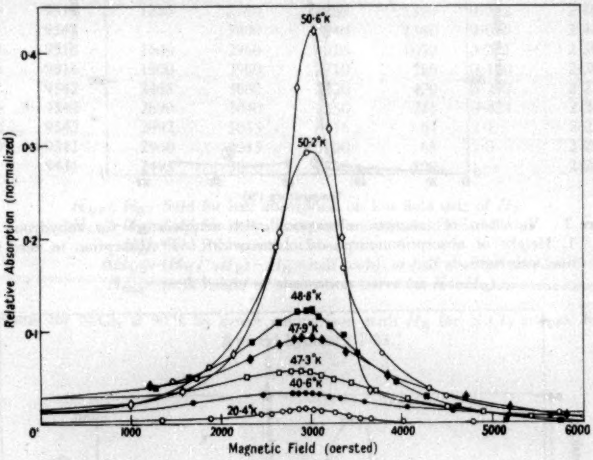


Figure 2. Magnetic resonance absorption curves for anhydrous $NiCl_2$ in the temperature range 20·4°K to 50·6°K.

Measurements were made on two different specimens of $NiCl_2$ and four different specimens of $CrCl_3$, though not all at the same temperature. The specimens varied in size, and results for different specimens were brought to the same scale of absorption by normalizing the peak relative absorption to 1·0 at 90°K for $NiCl_2$ and 20·4°K for $CrCl_3$, since measurements were made on all specimens at these temperatures. On plotting absorption curves from normalized results of different specimens, there was agreement between the plotted points thus justifying this procedure.

The results for $NiCl_2$ are presented in figures 2–5 and a summary of information extracted from the curves is given in table 2. Figure 2 shows the change in the resonance absorption between 20·4°K and 50·6°K, the absorption being shown relative to a maximum of 1·0 at 90·0°K. Between 50·6°K and 90°K the peak relative absorption rises to a maximum of about 2 at 60°K as is shown by additional data in table 2 and in figure 3. Figure 3 shows the variation in height of absorption curve, together with the variation of line width and zero field absorption, as a

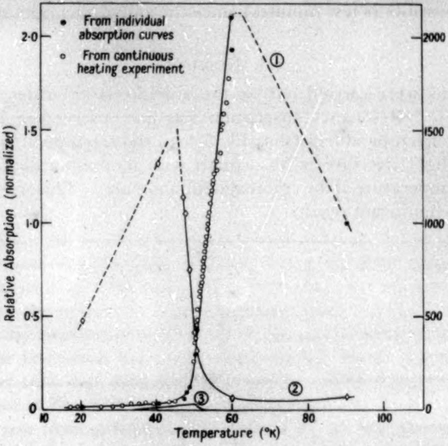


Figure 3. Variation of magnetic absorption with temperature for anhydrous $NiCl_2$. 1. Height of absorption curve. 2. Line width. 3. Absorption in zero applied magnetic field.

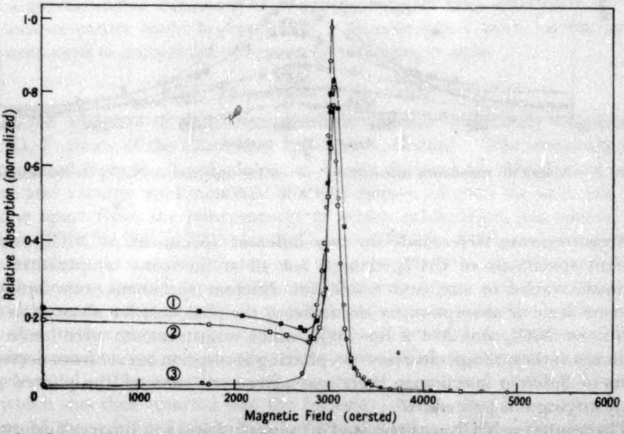


Figure 4. Magnetic resonance absorption in anhydrous $NiCl_2$ at 90°K. 1. On warming from 40·6°K. 2. On warming from 20·4°K. 3. On cooling from room temperature.

function of temperature. The curves of figure 4 were all obtained at 90°K but the two with large zero field absorption were observed after the specimen had been cooled below T_λ. In the latter case the specimen had been subjected to a field of 5500 oersteds at low temperature during the plotting of an absorption curve.

This type of curve was not found after the specimen had been in a high magnetic field at 90°K, a subsequent series of observations then leading to the normal paramagnetic resonance curve obtained on cooling the specimen to 90°K from room temperature. One such curve is shown in figure 4 for comparison. The zero field absorption as a function of temperature is shown on a larger scale in figure 5.

Table 2. Summary of Results for $NiCl_2$

T(°K)	ν(Mc/s)	$H_{1/2}<H_R$	H_R	$H_{1/2}>H_R$	$\Delta H_{1/2}$	A_{max}	$g=\dfrac{\nu}{1{\cdot}400\,H_R}$
20·4	9517	2280	2960	3435	475	0·015	2·30
40·6	9516	1220	2760	4080	1320	0·032	2·46
47·3	9542	—	2800	3940	1140	0·056	2·43
47·9	9516	1600	2960	4010	1050	0·091	2·30
48·8	9516	1900	2960	3710	750	0·120	2·30
50·2	9542	2465	3000	3420	420	0·292	2·27
50·6	9542	2690	3040	3350	315	0·424	2·24
60·0	9542	2992	3055	3116	61	2·1	2·23
90·0	9511	2960	3035	3100	65	1·0	2·24
293	9431	2495	3000	3590	590	—	2·25

$H_{1/2}<H_R$ = field for half absorption on low field side of H_R.
$H_{1/2}>H_R$ = field for half absorption on high field side of H_R.
H_R = field for maximum absorption.
$\Delta H_{1/2}=(H_{1/2}>H_R)-H_R$ = half width at half absorption.
A_{max} = peak height of absorption curve (at $H=H_R$)

g value for $NiCl_2$ at 90°K by direct comparison with H_R for $NiCl_2$ and αα diphenyl β picryl hydrazl = 2·23.

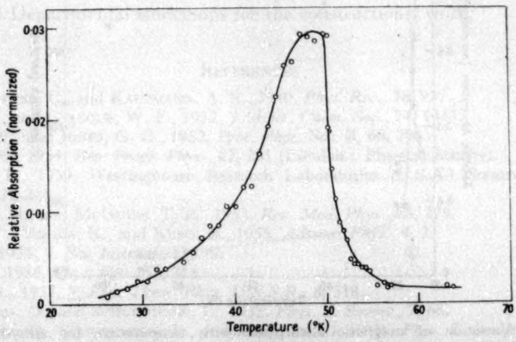


Figure 5. Variation of magnetic absorption in anhydrous $NiCl_2$ with temperature in zero magnetic field.

Results for $CrCl_3$ are presented in figures 6 and 7; figure 6 is analogous to figure 2 for $NiCl_2$ and figure 7 to figure 3 for $NiCl_2$. A summary of information derived from the curves is given in table 3.

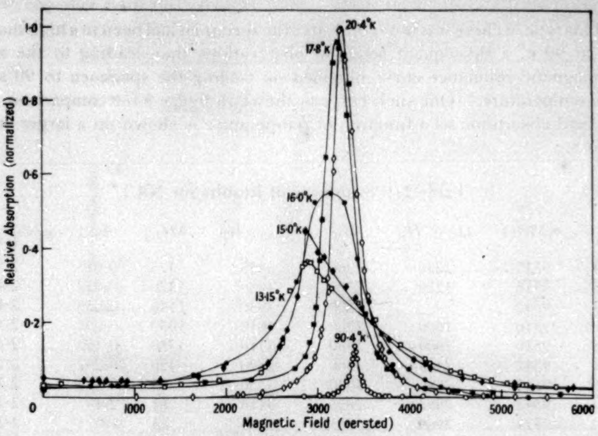


Figure 6. Magnetic resonance absorption curves for anhydrous $CrCl_3$ in the temperature range 13·15°K to 20·4°K.

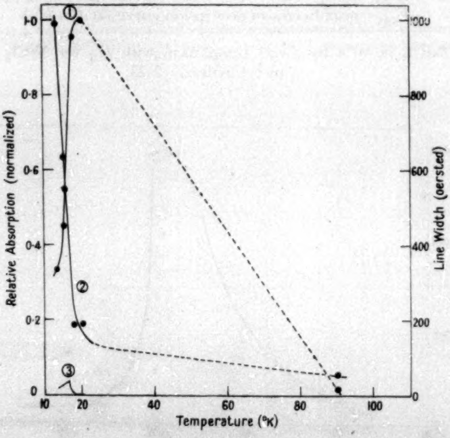


Figure 7. Variation of magnetic absorption with temperature for anhydrous $CrCl_3$. 1. Height of absorption curve. 2. Line width. 3. Absorption in zero applied magnetic field.

Table 3. Summary of Results for $CrCl_3$

T(°K)	ν(Mc/s)	$H_{1/2}<H_R$	H_R	$H_{1/2}>H_R$	$\Delta H_{1/2}$	A_{max}	$g=\dfrac{\nu}{1{\cdot}400\,H_R}$
13·15	9547	2665	2880	3870	990	0·370	2·37
15·0	9547	2720	2880	3510	630	0·450	2·30
16·0	9547	2770	2960	3510	550	0·550	2·30
17·8	9518	3040	3230	3415	185	0·955	2·11
20·4	9547	3155	3270	3460	190	1·00	2·09
90·0	9509	3352	3400	3454	54	0·14	2·00
293·0	9477	3340	3390	3454	62	0·0278	2·00

The table headings are as for table 2.
g value by direct comparison of H_R for $CrCl_3$ with H_R for αα diphenyl β picryl hydrazl at 90°K$=2{\cdot}00$.

§ 4. DISCUSSION OF RESULTS

The graphs of experimental results for $NiCl_2$ and $CrCl_3$ both show marked resemblance to those of other workers for antiferromagnetic substances (Maxwell and McGuire 1953).

Whilst this would be acceptable in the case of $NiCl_2$, $CrCl_3$ is not normally regarded as antiferromagnetic. Since there are important differences between the magnetic properties of these substances and those of normal antiferromagnets, and as the crystal structures do not lend themselves to antiferromagnetic ordering, it is worth while re-examining the whole question of the properties of $CoCl_2$, $FeCl_2$, $NiCl_2$ and $FeCl_3$ on the experimental evidence available. This will be done in part II.

ACKNOWLEDGMENTS

We wish to acknowledge the facilities made available to us by Professor G. O. Jones who first suggested the experimental investigation. We are indebted to the Ministry of Education for a maintenance allowance to one of us and for a grant-in-aid from the Central Research Fund of the University of London for the loan of apparatus.

Our thanks are also due to Mr. W. Doy and Mr. W. A. G. Baldock of the College and Departmental workshops for the constructional work.

REFERENCES

BITTER, F., STARR, C., and KAUFMANN, A. R., 1940, *Phys. Rev.*, **78,** 977.
BUSEY, R. H., and GIAUQUE, W. F., 1952, *J. Amer. Chem. Soc.*, **74,** 4443.
CHESTER, P. F., and JONES, G. O., 1953, *Proc. Phys. Soc.* B, **66,** 296.
LIDIARD, A. B., 1954, *Rep. Progr. Phys.*, **17,** 201 (London : Physical Society).
LINDER, C. T., 1950, Westinghouse Research Laboratories (U.S.A.) *Research Report*, R-94433-2-A.
MAXWELL, L. R., and MCGUIRE, T. R., 1953, *Rev. Mod. Phys.*, **25,** 279.
NAGAMIYA, T., YOSIDA, K., and KUBO, R., 1955, *Advanc. Phys.*, **4,** 2.
POTTER, H., 1934, *J. Sci. Instrum.*, **11,** 95.
ROSE, M. E., 1938, *Phys. Rev.*, **53,** 715.
SHALYT, S. S., 1938, *J. Exp. Theor. Phys. U.S.S.R.*, **8,** 518.
TRAPEZNIKOWA, O., and SCHUBNIKOW, L., 1935, *Phys. Z. Sowjet.*, **7,** 66.
TRAPEZNIKOWA, O., SCHUBNIKOW, L., and MILJUTIN, G., 1936, *Phys. Z. Sowjet.*, **9,** 237.
WOLTJER, H. J., 1925, *Leiden Comm.*, 173 b.
WOLTJER, H. J., and WIERSMA, E. G., 1929, *Leiden Comm.* 201 a.

Magnetic Properties of some Anhydrous Chlorides of the Iron Group

II : Interpretation of the Cryomagnetic Anomalies in $CoCl_2$, $FeCl_2$, $NiCl_2$ and $CrCl_3$

By J. W. LEECH and A. J. MANUEL†
Department of Physics, Queen Mary College, London

MS. received 1st *September* 1955

Abstract. It is shown that experimental evidence of magnetic properties and crystalline structure of $CoCl_2$, $FeCl_2$, $NiCl_2$ and $CrCl_3$ conflict with the assumption that these salts are antiferromagnetic. A two-dimensional ferromagnetic structure is indicated and the observed magnetization curves can be explained as due to the influence of high shape anisotropy associated with single layer domains.

§ 1. Introduction

The anomalies in the magnetic properties of $CoCl_2$, $FeCl_2$, $NiCl_2$ and $CrCl_3$ are associated with a peak in the specific heat curve of the type associated with magnetic ordering in ferromagnetic and antiferromagnetic substances.

The magnetic anomalies below the temperature T_λ at which the peak in the specific heat curve occurs are a variation of susceptibility with field strength and the existence of small remanences after magnetization in large magnetic fields.

The first experiments on these substances were carried out by Woltjer (1925) and later by Woltjer and Kamerlingh Onnes (1925) and Woltjer and Wiersma (1929). Further measurements have been made by Shalyt (1938), de Haas and Schultz (1939), de Haas, Schultz and Koolhaas (1940) and by Starr, Bitter and Kaufmann (1940). The magnetization curves obtained by the latter workers are shown in figure 1 and results of other workers for $CrCl_3$ are plotted in figure 2.

Landau (1933) pointed out that all the salts formed layer lattices, and proposed a magnetic structure of ferromagnetically ordered layers with the magnetic moments of alternate layers orientated in opposite directions. Schultz (1940) developed this hypothesis by postulating antiferromagnetic coupling between adjacent ferromagnetically ordered layers to account for the alternation of direction of orientation of the moments in the layers.

The recent interest in work on antiferromagnetic substances has led to $CrCl_2$, $FeCl_2$ and $NiCl_2$ being classified as antiferromagnetic on account of the maximum of zero field susceptibility near T_λ. The magnetic properties show however important differences from those of antiferromagnetics and the crystal structures are not compatible with antiferromagnetic ordering. It will be shown that both experimental and theoretical considerations are much more favourable to a form of ferromagnetic ordering.

---

† Now at Bedford College, London.

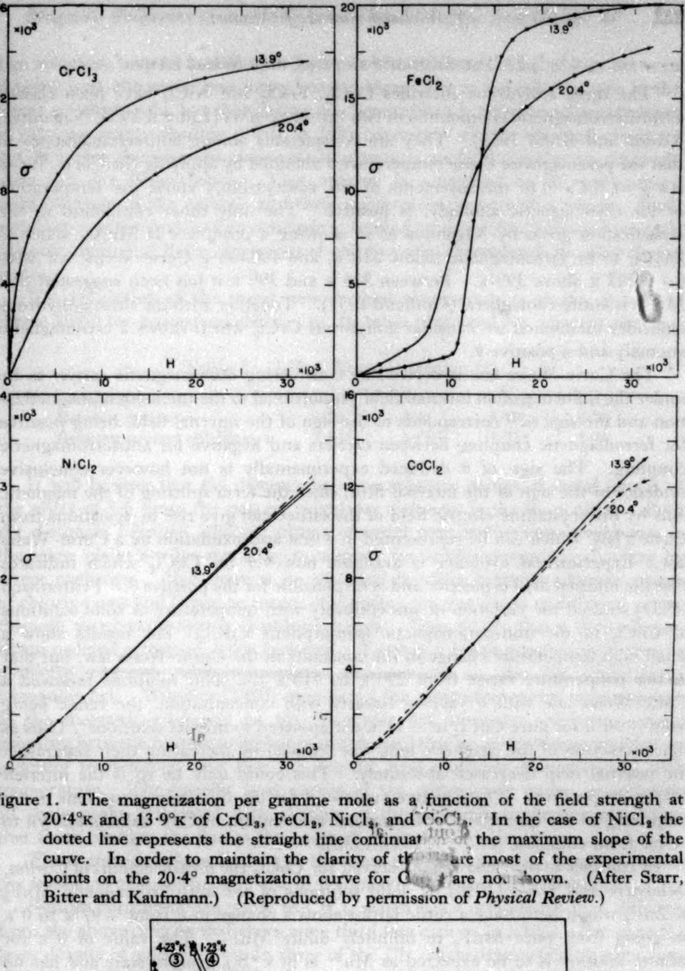


Figure 1. The magnetization per gramme mole as a function of the field strength at 20·4°K and 13·9°K of $CrCl_3$, $FeCl_2$, $NiCl_2$, and $CoCl_2$. In the case of $NiCl_2$ the dotted line represents the straight line continuation of the maximum slope of the curve. In order to maintain the clarity of the figure most of the experimental points on the 20·4° magnetization curve for $CoCl_2$ are not shown. (After Starr, Bitter and Kaufmann.) (Reproduced by permission of *Physical Review*.)

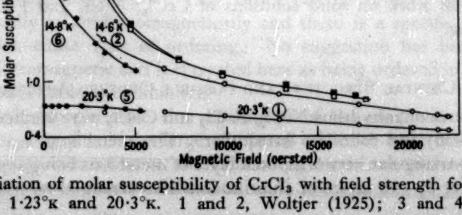


Figure 2. Variation of molar susceptibility of $CrCl_3$ with field strength for temperatures between 1·23°K and 20·3°K. 1 and 2, Woltjer (1925); 3 and 4, Woltjer and Kamerlingh Onnes (1925); 5 and 6, de Haas, Schultz and Koolhaas (1940).

§ 2. THE SIGN OF θ IN THE CURIE–WEISS LAW

The three anhydrous chlorides $CoCl_2$, $FeCl_2$ and $NiCl_2$ have been classed as antiferromagnetic compounds in two recent reviews (Lidiard 1954, Nagamiya, Yosida and Kubo 1955). They are conspicuous among antiferromagnetics in that the paramagnetic Curie temperature θ obtained by applying the Curie–Weiss law $\chi = C/(T-\theta)$ to measurements of the susceptibility above the temperature of the cryomagnetic anomaly, is positive. The only other compound in the classification given by Nagamiya *et al.* showing a positive θ is MnAs, which is known to be ferromagnetic below 318°K, and follows a Curie–Weiss law with $\theta = +293$°K above 399°K. Between 318°K and 399°K it has been suggested that MnAs is antiferromagnetic (Guillaud 1951). Together with the three anhydrous chlorides mentioned we consider anhydrous $CrCl_3$ which shows a cryomagnetic anomaly and a positive θ.

The Curie–Weiss law is derived by considering each magnetic carrier to be under the influence of an internal field proportional to the intensity of magnetization and the sign of θ corresponds to the sign of the internal field, being positive for ferromagnetic coupling between carriers and negative for antiferromagnetic coupling. The sign of θ deduced experimentally is not however conclusive evidence of the sign of the internal field, since the term splitting of the magnetic ions by the crystalline electric field of the lattice can give rise to deviations from Curie's law, which can be represented to a first approximation by a Curie–Weiss law. Experimental evidence is available however for $CoCl_2$ which indicates that the internal field is positive and is responsible for the positive θ. Fehrenbach (1937) studied the variation of susceptibility with temperature of solid solutions of $CoCl_2$ in the non-paramagnetic isomorphous $CdCl_2$. His results show a small high temperature change in the constants in the Curie–Weiss law, but that in the temperature range from 290°K to 410°K the solid solutions followed a Curie–Weiss law with θ varying linearly with concentration, the range being from $+30$°K for pure $CoCl_2$ to -17°K extrapolated to infinite dilution. Thus as the interaction of the magnetic ions was reduced by increasing their separation the internal field decreased absolutely. This could only be so if the internal field were positive. The negative value of θ for the infinitely dilute ions is not surprising as the quenching of the orbital momentum in Co^{2+} salts is known to be far from complete (cf. Van Vleck 1932).

In contrast to the behaviour observed for $CoCl_2$ the measurements of Corliss, Delabarre and Elliot (1950) on solid solutions of the antiferromagnetic MnF_2 in ZnF_2 which both have a rutile lattice show a change in θ from -97°K to 0°K on going from pure MnF_2 to infinitely dilute MnF_2. The value of 0°K for infinite dilution is to be expected as Mn^{2+} is in a $^6S_{5/2}$ ground state and has no resultant orbital moment.

Experimental work on solid solutions of $CrCl_3$, $NiCl_2$ and $FeCl_2$ is unfortunately lacking.

§ 3. THE CRYSTAL STRUCTURE AND POSSIBLE ORDERED ARRANGEMENTS

The structures of anhydrous $NiCl_2$, $FeCl_2$ and $CoCl_2$ were studied by Pauling and Hoard (1930) and found to be similar. The metal ions are arranged in planes forming triangular networks, each layer of metal ions being separated from an adjacent layer by two layers of chlorine ions also arranged on triangular nets.

The structure may be looked upon as a pile of sandwiches of metal ions between chlorine ions, the metal ions being bonded chemically to the chlorine ions in the planes on either side, but there being no chemical bonds between adjacent planes of chlorine ions. Pauling (1940, p. 189) comments on the identical structure of $CdCl_2$ that the adjacent planes of chlorine ions are held together by relatively weak van der Waals forces. This opinion is supported by the pronounced basal cleavage of the crystals and their sublimation in very thin plates. Each metal ion is at the centre of an octahedron formed by three chlorine ions from each adjacent layer, and the arrangement of chlorine ions corresponds to cubic closest packing.

The separation of the metal ions in these compounds is shown in table 1 (Starr 1940) for nearest neighbours in the same plane and for adjacent planes.

Table 1. Metal Atom Distances (Å)

Compound	$CoCl_2$	$FeCl_2$	$NiCl_2$	$CrCl_3$
Between layers	5·79	5·84	5·78	5·76
Between nearest neighbours	3·546	3·580	3·544	3·46

It will be seen that the distance between adjacent planes of metal ions is of the order of 1·6 times the nearest neighbour distances in the same plane. Both the distances between adjacent planes of metal ions and nearest neighbours in the same plane are far too large to account for antiferromagnetic coupling by direct exchange. Also there is no chemical bonding between adjacent planes of chlorine ions and hence no means of coupling the metal ions in different planes through super-exchange or double-exchange. This means that a three-dimensional ordered structure is not favoured. Further it is not possible to produce a perfectly ordered antiferromagnetic arrangement in the two-dimensional triangular net. Wannier (1950) considered the minimum energy arrangements of the magnetic moments on the triangular net assuming antiferromagnetic coupling between nearest neighbours and showed that no such arrangement has an ordering temperature and that the specific heat increases monotonically with temperature. The specific heat curves of the compounds under examination all show peaks associated with changes in the magnetic properties suggesting that these changes are due to the onset of co-operative phenomena.

These considerations lead us to the view that the magnetic structure below the transition temperatures is one of ferromagnetic single layer domains.

Similar considerations apply to $CrCl_3$ which has a crystal structure differing from the above only in that there are a third less cations in a layer relative to the layer density of anions. One in three of the Cl^- octahedra containing metal ions are unoccupied in $CrCl_3$. The layer lattice of Cr^{3+} ions is a honeycomb net with three nearest neighbours. This lattice can be ordered perfectly either ferromagnetically or antiferromagnetically and there is a specific heat anomaly associated with either type of ordering. No suggestion has been made that $CrCl_3$ is antiferromagnetic and it is treated here as being ordered in the same way as the other salts under examination.

On the microscopic scale the separations of the magnetic ions fulfil the condition that they should be large compared with the ionic radius, to give a positive exchange interaction on the Heisenberg model, or narrow 3d bands on the collective electron theory of Stoner.

§ 4. MAGNITUDE OF THE INTERNAL FIELD AND MAGNETIZATION CURVES

We consider here the degree of magnetization obtained in high magnetic fields, which however are small compared with the equivalent internal field on the two-dimensional Ising model, and show that the magnetizations are larger than would be possible with antiferromagnetic coupling.

Wannier (1950) and Newell and Montroll (1953) give equivalent expressions for the ordering temperature of the triangular and hexagonal nets in terms of the exchange interaction between nearest neighbours. If $|J|$ is the magnitude of the exchange integral and $|J|\,\boldsymbol{\sigma}_1 . \boldsymbol{\sigma}_2$ represents the difference in energy between a parallel and antiparallel arrangement of two spins where $|\boldsymbol{\sigma}_1| = |\boldsymbol{\sigma}_2| = 1$ and if $K = |J|/kT_{\mathrm{c}}$ where T_{c} is the ordering temperature then $\sinh^2 K_{\mathrm{H}} = \frac{1}{2}$ for the hexagonal lattice, $4K_{\mathrm{T}} = \ln 3$ for the triangular lattice.

For ferromagnetic coupling $\boldsymbol{\sigma}_1 = \boldsymbol{\sigma}_2$ and for antiferromagnetic coupling $\boldsymbol{\sigma}_1 = -\boldsymbol{\sigma}_2$. An equivalent exchange field can be defined by $\mu H_j = z|J|$ where μ is the magnetic moment of an ion and z is the number of nearest neighbours. Then $H_j = zkT_{\mathrm{c}}K/\mu$. Using the values of T_{c} given by the specific heat anomalies and the spin only value of μ the equivalent exchange fields are obtained as in table 2.

Table 2

	Normal state	Magnetic moment spin only	T_{c}(°K)	H_j(Oe) × 10^{-5}
$\mathrm{CoCl_2}$ $\mathrm{Co^{2+}}$	$^4\mathrm{F}_{9/2}$	3·87β	24·9	1·58
$\mathrm{FeCl_2}$ $\mathrm{Fe^{2+}}$	$^5\mathrm{D}_4$	4·90β	23·5	1·177
$\mathrm{NiCl_2}$ $\mathrm{Ni^{2+}}$	$^3\mathrm{F}_4$	2·83β	49·6	4·25
$\mathrm{CrCl_3}$ $\mathrm{Cr^{3+}}$	$^4\mathrm{F}_{3/2}$	3·87β	16·8	1·33

If we compare the magnetization of these salts attained in fields of 30 000 oersteds (i.e. small compared with the equivalent exchange fields) by taking the fraction of a saturation magnetization corresponding to a spin only value of the magnetic moment, we find that the fraction of saturation attained is far too high to be consistent with an antiferromagnetic field of the order calculated. The comparison is shown in table 3 the magnetizations at 13·9°K being taken from the curves of Starr, Bitter and Kaufmann (1940) reproduced in figure 1.

Table 3

Compound	σ(30 000 Oe, 13·9°K)	σ_{sat} (spin only)	$\dfrac{\sigma}{\sigma_{\mathrm{sat}}}$
$\mathrm{CoCl_2}$	12 600	16 750	0·752
$\mathrm{FeCl_2}$	20 000	27 980	0·716
$\mathrm{NiCl_2}$	2 800	22 350	0·126
$\mathrm{CrCl_3}$	13 500	16 750	0·805

σ is in c.g.s. units per gm molecule.

The calculations of the equivalent exchange field and the saturation magnetization are based on simple assumptions and, on the Ising model of the triangular lattice, can only refer to ferromagnetic ordering. Should more refined calculations lead to a transition temperature associated with minimum energy arising from partial ordering of the triangular lattice where antiferromagnetic coupling is favoured between nearest neighbours one should expect an equivalent exchange field of the order calculated.

§ 5. The Magnetic Hardness

The preceding examination of the experimental evidence and the crystal structure suggests that the salts considered are ferromagnetic with single layer domains. The exceptional magnetic hardness may be attributed to large anisotropy. In addition to the usual causes of anisotropy, that due to the particular shape of the domains must be considered. Stoner and Wohlfarth (1948) have considered in detail the effect of shape anisotropy arising from differences in demagnetizing factors along different axes of ellipsoids of revolution and obtained magnetization curves for assemblies of randomly orientated identical ellipsoids. Their curves are shown in figure 3. The full curves are for assemblies of prolate spheroids in the absence of other forms of anisotropy when the axis of easy magnetization would be along the polar axis. The initial magnetization curve is only reversible as long as the spontaneous magnetization of domains is not reversed discontinuously. After application of high fields discontinuous reversals take place and the hysteresis curve is obtained. The broken curves of figure 3 are for assemblies of oblate spheroids with the directions of easy magnetization along the equatorial axes. The ordinates and abcissae in figure 3 are the reduced quantities $\cos\phi = I/I_0$ and $h = H/|N_a - N_b|I_0$ where H is the applied field, I the magnetization, I_0 the saturation magnetization and N_a and N_b the demagnetizing factors along and perpendicular to the polar axis. The form of the curves shows clear resemblance to the experimental curves for the anhydrous chlorides in figure 1, that for oblate spheroids corresponding to $CrCl_3$ and for prolate spheroids to the remainder.

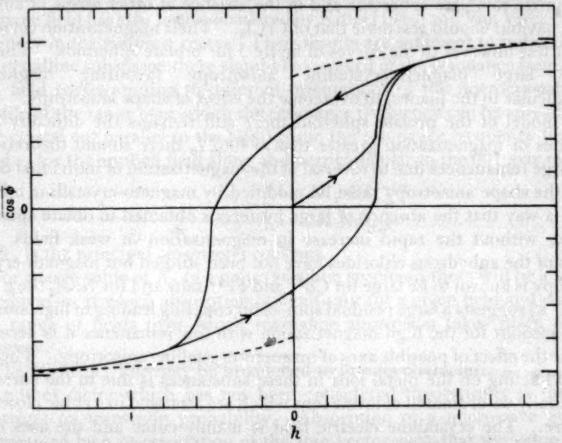


Figure 3. Magnetization curves for prolate (full curves) and oblate (broken curves) spheroids orientated at random. The mean resolved magnetization per spheroid in the positive field direction, or the resultant magnetization in this direction of the assembly, is proportional to $\overline{\cos\phi}$. $H = (|N_a - N_b|)I_0 h$. (*After* Stoner and Wohlfarth 1948, and Rhodes 1948.)

Single layer domains can be idealized as oblate spheroids and in the absence of other than shape anisotropy their easy directions of magnetization should lie in the planes. The magnetization curve for an assembly of such domains should resemble the broken curve in figure 3. Allowing for the departure from ideal disc shaped domains the behaviour of powdered specimens of $CrCl_3$ is explicable in this way, assuming that other forms of anisotropy are small. The causes of magneto-crystalline anisotropy are still uncertain but believed to be due to incomplete quenching of the orbital momenta in the crystalline field. The g value of 2·0 for $CrCl_3$ suggests that the orbital moment is completely quenched and magneto-crystalline anisotropy should be small. The cases of $CoCl_2$, $FeCl_2$, and $NiCl_2$ present difficulties on the idealized model of Stoner and Wohlfarth. They approximate to oblate spheroids and in the absence of other forms of anisotropy their behaviour should resemble that of $CrCl_3$. Their magnetization corresponds more or less however to the curves of figure 3 for prolate spheroids. This would require large magneto-crystalline anisotropy favouring magnetization perpendicular to the planes, to overcome the effect of shape anisotropy. For this simple model of the prolate spheroid there still remains the difficulty that for intensities of magnetization greater than $0{\cdot}4062\,I_0$ there should theoretically be quite large remanences due to reversal of the magnetization of individual domains. Hence the shape anisotropy must be modified by magneto-crystalline anisotropy in such a way that the absence of large hysteresis obtained in oblate spheroids is retained, without the rapid increase in magnetization in weak fields. Single crystals of the anhydrous chlorides have not been studied but magneto-crystalline anisotropy is known to be large for Co^{2+} and Fe^{2+} salts and for $NiCl_2$ the g value of 2·23 (90°K) suggests a large residual spin–orbit coupling leading to high anisotropy.

To account for the high magnetization with low remanence it is necessary to examine the effect of possible axes of magneto-crystalline anisotropy. The crystalline field acting on the metal ions in these substances is due to the surrounding octahedrally co-ordinated chlorine ions which are arranged in a cubic close-packed structure. The crystalline electric field is mainly cubic and the axes of cubic symmetry are the body diagonals of the chlorine octahedra.

There is a trigonal axis of symmetry normal to the plane of the metal ions. It can be seen that if the cubic axes are easy or difficult directions of magnetization on account of magneto-crystalline anisotropy, the oblate spheroid model would be modified so that to a crude approximation there would be three equivalent easy and difficult directions within the equatorial plane of the spheroid. Further an assembly of such domains could approach a high fraction of saturation without discontinuities more slowly than in the ideal case, while still retaining the reversible character of the magnetization curve for an assembly of ideal oblate spheroids.

The assumption that the layers of metal ions in the salts considered form single domains is justified on the demonstration of Kittel (1946) that for the thinnest films of ferromagnetic substances the lowest energy corresponds to the whole film being spontaneously magnetized as a single domain. Drigo and Pizzo (1948) observed that the Barkhausen effect disappeared in thin films and the recent work of Crittenden and Hoffman (1953) on the magnetization curves of thin ferromagnetic films bears out Kittel's theory.

§ 6. DISCUSSION OF MAGNETIC RESONANCE

The graphs of the experimental results for both $NiCl_2$ and $CrCl_3$ show a marked resemblance to those of other workers for antiferromagnetic substances (e.g. Maxwell and McGuire 1953). In the light of the discussion of the interpretation of the cryomagnetic anomalies in § 4 (and that $CrCl_3$ which is not classed as an antiferromagnetic substance shows similar behaviour) it is necessary to examine whether the results obtained are consistent with the ferromagnetic structure suggested in the preceding section.

The experimental conditions differ from those which have been assumed in the theoretical treatment of ferromagnetic resonance. Kittel (1948) in his macroscopic theory assumes that the material is magnetized to saturation in the direction of the applied field and Van Vleck (1950) makes equivalent assumptions in his quantum mechanical treatment. These conditions certainly do not apply to the present work, where the material is extremely anisotropic and far from saturation. The effect of anisotropy on the ferromagnetic resonance frequency for a given field has also been considered by Kittel (1948) and Van Vleck (1950), but again only in the simplest cases. Their results are sufficient to indicate that in a polycrystalline substance there should be a spread of the resonance frequency for a given field corresponding to different orientations of the microcrystals. As an example of the range over which the resonance frequency can be spread for a thin cubic crystal cut parallel to the (100) plane the resonance frequency lies between ν_1 and ν_2 for the applied field along and perpendicular to the 001 axis respectively where

$$\frac{\nu_1}{\nu_2}=\left[\frac{1+2K_1/HI_s}{1-2K_1/HI_s}\right]^{1/2}$$

and K_1 is the principal anisotropy constant.

For a large value of K_1 in a polycrystalline specimen there will be a large range of frequencies at which absorption is significant for a given field and conversely a large range of fields over which resonance absorption takes place at a single frequency.

The resonance line may be broadened to lose its character.

Another factor which has not been considered theoretically is the effect of anisotropy on transition probability. Absorption of a microwave quantum is accompanied by a re-orientation of the spin vector such that the value of M_s the component of spin momentum along the axis of quantization in units of $h/2\pi$ changes by ± 1. Apart from requiring different energy on account of anisotropy it may be that a direction of hard magnetization lies between the initial and final orientations of the spin vector and the transition probability is reduced by the energy barrier between the two states. This is another possible explanation of the reduction in absorption when the crystals are highly anisotropic.

The above considerations show that the observed variation of absorption with temperature is consistent with the interpretation of $NiCl_2$ and $CrCl_3$ as highly anisotropic ferromagnetic substances.

For $CrCl_3$ there is further evidence in the absorption curves of ferromagnetic properties. Figure 6 of part I shows that below the transition temperature indicated by the specific heat anomaly, the resonance curves are asymmetrical. The increase in absorption towards the peak of the curves is more rapid than the decline on the high field side. Bearing in mind that these curves are probably the superposition of a number of resonance curves there is evidence of the existence of an anti-resonance point such as has been found in ferromagnetic substances (Yager 1949), and discussed by Kittel (1951). The field $H_{\min}$ at which the anti-resonance point occurs is given by Kittel as $H_{\min}=H_R/(1+4\pi\chi_v)$ where H_R is the field for maximum absorption and χ_v the volume susceptibility. Taking the value $\chi_{\mathrm{mol}} \sim 1{\cdot}5$ (de Haas, Schultz and Koolhaas 1940) and $H_R = 2880$ oersteds for the curve at 13·15°K we find $H_{\min} = 2180$ oersteds. This is of the required order to account for anti-resonance producing the observed curve.

The fact that the absorption curves of figure 4 of part I for $NiCl_2$ show large zero field absorption at 90°K after cooling the salt below the transition temperature suggests a partial persistence of the ordered state of the magnetic carriers above the transition temperature which is broken up by the application of a field of the order of 2800 oersteds. The large absorption at zero applied field could be due to resonant absorption by some of the ions in an inhomogeneous internal field due to a metastable ordered state remaining after cooling below the transition temperature. There appears to be thermal hysteresis in the magnetic transition.

Thermal hysteresis was reported by Trapeznikowa, Schubnikow and Miljutin (1936) in their specific heat measurements on $NiCl_2$. Their results show two peaks in the specific heat curve, the major one at 49·6°K and a smaller one at about 57°K. The position of the 57°K peak varied depending on the temperature to which the specimen had been cooled before the specific heat measurements were made. Busey and Giauque (1952) find only one anomaly at 52·35°K but all their measurements were made after the specimen had been cooled to liquid hydrogen temperature. An examination of the results of Trapeznikowa *et al.* shows that their final table is a composite one made up of measurements taken over a large range of temperatures with different starting temperatures. If there is thermal hysteresis associated with the transition in $NiCl_2$ the multiple anomaly found by Trapeznikowa *et al.* is explained by the superposition of specific heat curves taken with different pre-cooling temperatures. The absence of this effect in the measurements of Busey and Giauque may be because their starting temperature was always well below the transition temperature.

Busey and Giauque suggest that the difference between their results and those of Trapeznikowa *et al.* may be due to thermal lag in the latter's glass calorimeter but the peak at 57°K was obtained when the specimen was not cooled below 54°K.

The variation of magnetic absorption with temperature in the absence of an external field shown in figure 5 of part I may be interpreted as resonance by a fraction of the magnetic ions in the inhomogeneous internal field present during the break up of the magnetic ordering.

The effective values of the spectroscopic splitting factor g quoted in tables 2 and 3 have not been corrected for the demagnetizing factors as these are not readily calculated. This is not important for $NiCl_2$ as the magnetization of the specimen is relatively small but in $CrCl_3$ the increase in effective g value, rising to about 15% at 13·15°K, corresponds to what would be expected for a thin specimen as was used in the experiments.

The line shape for paramagnetic resonance in $NiCl_2$ and $CrCl_3$ indicates strong exchange interactions on Van Vleck's (1948) theory of exchange narrowing in paramagnetic resonance. Van Vleck has given numerical examples of the effect of exchange interaction in cubic crystals. For no exchange interaction, assuming other forms of broadening lead to a gaussian line shape, the ratio

$$\frac{[\langle \Delta H^4 \rangle_{\mathrm{av}}]^{1/4}}{[\langle \Delta H^2 \rangle_{\mathrm{av}}]^{1/2}} = 1{\cdot}32$$

where
$$[\langle \Delta H^n \rangle_{\mathrm{av}}]^{1/n} = \left[\frac{\int_0^\infty A(H)\,[H - H_R]^n\, dH}{\int_0^\infty A(H)\, dH} \right]^{1/n};$$

$A(H)$ is the relative absorption in applied field H and H_R the magnetic field corresponding to maximum absorption. If exchange interaction is present then the ratio

$$\frac{[\langle \Delta H^4 \rangle_{\text{av}}]^{1/4}}{[\langle \Delta H^2 \rangle_{\text{av}}]^{1/2}} > 1{\cdot}32 .$$

For $CrCl_3$ at 90°K this ratio is 2·2 and for $NiCl_2$ at the same temperature it is 1·74. Although the crystal structure is not by any means cubic these large values indicate in both cases the importance of exchange narrowing. The calculations were carried out by summing numerically for values of $dH = 10$ oersteds.

Finally, in addition to the low temperature measurements, room temperature measurements were made to compare with those of other workers.

$CrCl_3$ has been examined by Bagguley *et al.* (1948), Ting and Williams (1951) and Ting, Farringer and Williams (1955). The former workers quoted a half-width at half peak absorption of about 50 oersteds. Ting and Williams (1951) obtained a width at half peak absorption of 610 oersteds (which they call a half-width) and in 1955 a width of 148 oersteds corresponding to a half-width at half peak absorption of 74 oersteds. The present result of 62 oersteds is an exact mean of the other two.

$NiCl_2$ was examined by Ting and Williams (1951) when they obtained a width at half absorption of 2020 oersteds. This is about double the value of 1180 oersteds obtained at room temperature in the present measurements. The high value of line width obtained for $CrCl_3$ in the same series of measurements by Ting and Williams suggests that those for $NiCl_2$ are also erroneous. Their 1955 measurements were carried out on fresh samples of $CrCl_3$ in a different spectrometer. It cannot be inferred whether their earlier measurements were in error due to the presence of water molecules in the salts or some defect in experimental procedure such as failure to retune the cavity to correct for dispersion in regions of high absorption.

The g values at room temperature in this investigation are in agreement with those of the other workers.

§ 7. CONCLUSION

The experimental evidence of the magnetic, thermal and crystalline structure of the anhydrous chlorides $CoCl_2$, $FeCl_2$, $NiCl_2$, $CrCl_3$ together with the variation of paramagnetic resonance absorption in $CrCl_3$ and $NiCl_2$ suggest that these salts become ferromagnetic at low temperatures, their exceptional magnetic hardness and the similarity of the magnetic behaviour of the first three to antiferromagnetic compounds being due to a two-dimensional domain structure. This two-dimensional nature of the domains makes domain boundaries energetically unfavourable and the assembly of single domains gives rise to the great magnetic hardness. The magnetic behaviour is found to be consistent with the theory of magnetization of an assembly of single domain particles of Stoner and Wohlfarth (1948).

ACKNOWLEDGMENTS

We wish to acknowledge the facilities made available to us by Professor G. O. Jones who first suggested the experimental investigation. We are indebted to the Ministry of Education for a maintenance allowance to one of us and for a grant-in-aid from the Central Research Fund of the University of London for the loan of apparatus.

Our thanks are also due to Mr. W. Doy and Mr. W. A. G. Baldock of the College and Departmental workshops for the constructional work.

REFERENCES

BAGGULEY, D. M. S., BLEANEY, B., GRIFFITHS, J. H. E., PENROSE, R., and PLUMPTON, I., 1948, *Proc. Phys. Soc.*, **61**, 542.
BUSEY, R. H., and GIAUQUE, W. F., 1952, *J. Amer. Chem. Soc.*, **74**, 4443.
CORLISS, L., DELABARRE, Y., and ELLIOT, N., 1950, *J. Chem. Phys.*, **18**, 1256.
CRITTENDEN, E. C., and HOFFMAN, R. W., 1953, *Rev. Mod. Phys.*, **25**, 310.
DRIGO, A., and PIZZO, M., 1948, *Nuovo Cim.*, **5**, 1.
FEHRENBACH, C., 1937, *J. Phys. Radium*, **8**, 11.
GUILLAUD, C., 1951, *J. Phys. Radium*, **12**, 223.
DE HAAS, W. J., and SCHULTZ, B. H., 1939, *Leiden Comm.*, 256 d.
DE HAAS, W. J., SCHULTZ, B. H., and KOOLHAAS, J., 1940, *Leiden Comm.*, 259 a.
KITTEL, C., 1946, *Phys. Rev.*, **70**, 965; 1948, *Ibid.*, **73**, 155; 1951, *J. Phys. Radium*, **12**, 291.
LANDAU, L., 1933, *Phys. Z. Sowjet.*, **4**, 675.
LIDIARD, A. B., 1954, *Rep. Progr. Phys.* **17**, 201 (London: Physical Society).
MAXWELL, L. R., and MCGUIRE, T. R., 1953, *Rev. Mod. Phys.*, **25**, 279.
NAGAMIYA, T., YOSIDA, K., and KUBO, R., 1955, *Advanc. Phys.* **4**, 2.
NEWELL, G. F., and MONTROLL, E. W., 1953, *Rev. Mod. Phys.*, **25**, 353.
PAULING, L., 1940, *The Nature of the Chemical Bond* (Oxford: University Press).
PAULING, L., and HOARD, J., 1930, *Z. Kristallogr.*, **74**, 546.
SCHULTZ, B. H., 1940, *Leiden Comm.*, 259 b.
SHALYT, S. S., 1938, *J. Exp. Theor. Phys. U.S.S.R.*, **8**, 518.
STARR, C., 1940, *Phys. Rev.*, **58**, 984.
STARR, C., BITTER, F., and KAUFMANN, A. R., 1940, *Phys. Rev.*, **58**, 977.
STONER, E. C., and WOHLFARTH, E. P., 1948, *Phil. Trans. Roy. Soc.* A, **240**, 599.
TING, Y., FAIRRINGER, L. F., and WILLIAMS, D., 1955, *Phys. Rev.*, **97**, 1037.
TING, Y., and WILLIAMS, D., 1951, *Phys. Rev.*, **82**, 507.
TRAPEZNIKOWA, O., SCHUBNIKOW, L. F., and MILJUTIN, G., 1936, *Phys. Z. Sowjet.*, **9**, 237.
VAN VLECK, J. H., 1932, *Phys. Rev.*, **41**, 208; 1948, *Ibid.*, **74**, 1168; 1950, *Ibid.*, **78**, 266.
WANNIER, G. H., 1950, *Phys. Rev.*, **79**, 357.
WOLTJER, H. J., 1925, *Leiden Comm.*, 173 b.
WOLTJER, H. J., and KAMERLINGH ONNES, H., 1925, *Leiden Comm.*, 173 c.
WOLTJER, H. J., and WIERSMA, E. G., 1929, *Leiden Comm.*, 201 a.
WOOSTER, N., 1930, *Z. Kristallogr.* **74**, 363.
YAGER, W. A., 1949, *Phys. Rev.*, **75**, 316.

Measurement of Noise Spectra of a Germanium p–n Junction Diode

By F. J. HYDE
Radio Research Station, Slough, Bucks.

Communicated by R. L. Smith-Rose; MS. received 21st September 1955 and in final form 14th November 1955

Abstract. Measurements have been made of the excess noise generated in a fused-alloy type p–n junction diode in the frequency range 0·12 c/s to 2 Mc/s at 29°c with the reverse direct current I as parameter. The observed noise spectral density may be synthesized from three well-defined types of component: (*a*) an extensive component proportional to f^{-1} directly observed over five decades, and by synthesis assumed to exist over as many as seven decades of frequency; (*b*) a component proportional to $(1+\omega^2\tau_1^2)^{-1}$ associated with a single relaxation time τ_1; (*c*) a uniform component associated with the 'shot' noise of the measured current I, for frequencies less than $1/2\pi\tau_\mathrm{p}$ where τ_p is the hole lifetime. τ_1 was found to increase with increasing I while the intensity of the f^{-1} component increased more rapidly than in proportion to I^2.

§ 1. INTRODUCTION

Detailed measurements of the noise, in *excess* of the thermal noise associated with the real part of the internal admittance of a point-contact germanium diode biased in the reverse direction have already been published by the author (Hyde 1953, hereafter to be referred to as I). The present series of measurements also refer to the excess noise spectrum of a commercial fused-alloy type junction diode, made from n-type material, as a function of reverse bias, with the ambient temperature maintained constant. Measurements have been made in terms of a noise current generator represented by a source of noise current in parallel with the a.c. terminal admittance of the diode. The noise was measured as the mean square noise current $\langle\Delta i^2\rangle$† in a noise bandwidth Δf and is presented in the paper as the spectral density of mean square noise current $S_\mathrm{I}(f)=\langle\Delta i^2\rangle/\Delta f$ amp² sec.

It was found that the selection of a junction diode, which was a stable noise source over a wide range of reverse bias, presented a similar problem to that encountered in seeking a point-contact diode. It is the exception rather than the rule that diodes, biased in the reverse direction, show noise that is steady. More generally, the noise may occupy different levels in turn and switch in an apparently random manner between them. The duty cycle, when switching occurs, is usually short—a few seconds or minutes—so that it is difficult to separate the levels at low frequencies.

There seem to be no indications on the macroscopic scale (i.e. from measurements of d.c. characteristics or impedance) that a given diode will show more than one level of noise. The noise from the present specimen showed a single level except in the neighbourhood of 100 μA reverse current, for which three different levels were observed in the frequency range between about 300 c/s and 100 kc/s. The points in figure 4 refer to the lowest or 'basic' level, which was the most prevalent one, and in fact, for most of the time, the only one. The other levels are not shown because it was not possible to resolve them accurately; by monitoring the amplified noise on an oscilloscope, however, it was noted that they did not represent an increase of more than 3 dB above the basic level at any frequency of measurement.

---

† Throughout the paper ⟨ ⟩ means 'mean value of'.

The possibility of switching, with a short duty cycle, having occurred at measuring frequencies below 300 c/s cannot be excluded. It would have been obscured by the statistical variations of the noise itself, which become more evident as the measuring frequency is reduced, through the smoothing circuits used in the measuring equipment becoming less effective. For this reason, if switching did occur below 300 c/s for this particular reverse current (100 μA), noise readings would be greater than those to be associated with the basic level. The subsequent analysis of the noise spectrum for $I = 100\,\mu$A shows that there is evidence in favour of this.

§ 2. EXPERIMENTAL METHOD

The amplifiers used were as in I. The ambient temperature T of the diode was maintained at $29 \pm 0{\cdot}01$°C in an oil bath, with a very low drift rate between the extremes. Temperature variations were measured by observing the time variation of resistance of a thermistor in the bath. Excluding long-term drift between the extremes mentioned it was found that the r.m.s. value of temperature fluctuations was of the order of 0·001°C or less.

After selective amplification the noise was rectified using a square-law detector. For a typical measurement a reading proportional to the excess noise was obtained as a difference between two deflections: the first due to the total noise originating in the diode and its associated feed circuit and in the amplifier; the second due to the 'background' thermal noise originating in the feed circuit, the real part of the diode admittance and the equivalent noise resistance of the amplifier. To obtain this latter reading it was necessary to replace the diode in the feed circuit by an admittance comprised of passive resistance R_T and capacitance C_T components equal to the measured diode admittance at the frequency in question.

At high frequencies and low currents the impedance of the diode became small (of the order of 100 ohms). Under these same conditions the excess noise was also small (as low as 0·2 times the thermal noise in 100 ohms at 29°C) so that the noise voltage developed at the terminals of the diode could not be detected above the receiver noise if direct connection was made to the receiver. A coupling circuit which was used to give an improvement in the 'signal' to noise ratio is described in Appendix I.

The components R_T and C_T were measured in a.c. bridge circuits. In view of the dependence of R_T and C_T upon the applied voltage at the working frequency, this voltage was chosen as small as possible so that the true 'small-signal' parameters were measured: that this was the case was assured by reducing the amplitude of the signal voltage in steps until further reduction produced no change of bridge balance. The variation of R_T and C_T with frequency are shown in figures 1(a) and 1(b) with the reverse current I as parameter. An analysis of these data is described in Appendix II. The analysis leads to the equivalent circuits of figures 2(a) and 2(b) for I less or greater than 60 μA respectively. The

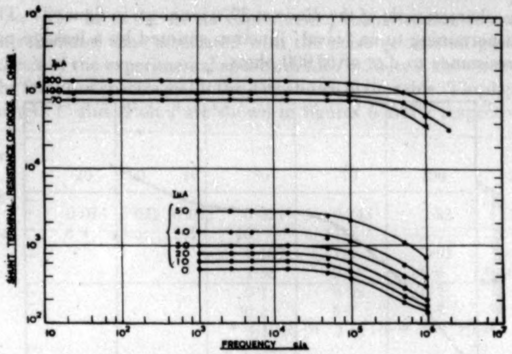


Figure 1(*a*). Shunt terminal resistance of diode at 29°c with reverse current as parameter.

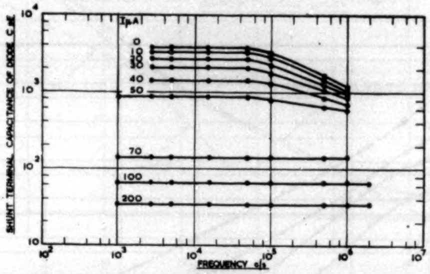


Figure 1(*b*). Shunt terminal capacitance of diode at 29°c with reverse current as parameter.

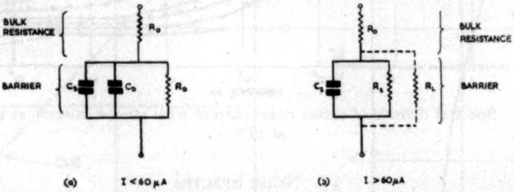

BULK RESISTANCE
BARRIER
C_S
C_D
R_0
R_B
(a)
I < 60 μA
R_0
C_S
R_S
R_L
BULK RESISTANCE
BARRIER
(b)
I > 60 μA

Figure 2. Small-signal equivalent circuit for diode for (*a*) $I < 60\,\mu\text{A}$, (*b*) $I > 60\,\mu\text{A}$.

most important parameters deduced from the analysis are the lifetime τ_p of holes in the junction ($3{\cdot}1\,\mu$sec), the zero-voltage low-frequency ($\omega\tau_p \ll 1$) junction resistance R_{D0} (450 ohms) and the resistance R_0 of the bulk germanium in series with the junction (50 ohms).

The d.c. characteristic of the diode at 29°c is shown in figure 3. This may be analysed as pertaining to an 'ideal' junction shunted by a leakage path having an ohmic resistance to d.c. of 68 000 ohms.

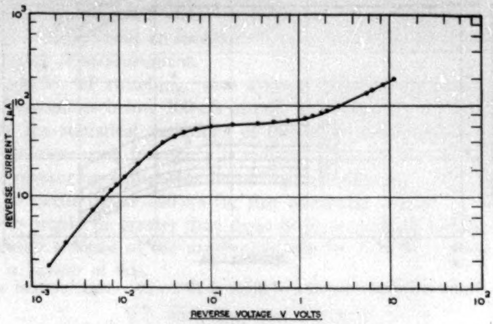


Figure 3. D.C. characteristic of diode at 29°c.

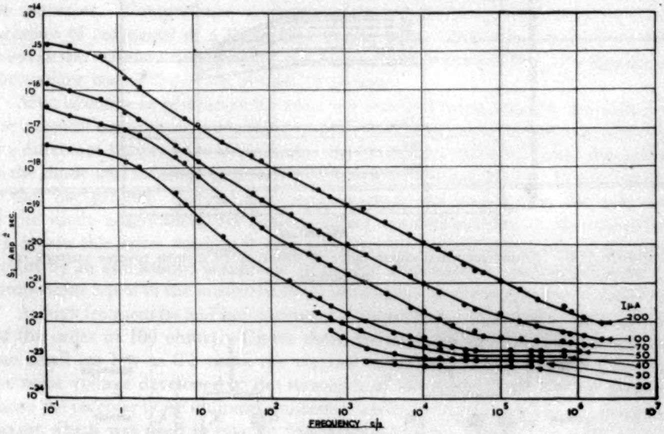


Figure 4. Spectral density of excess noise current with reverse current as parameter at 29°c.

§ 3. NOISE SPECTRA

The spectral density of excess noise current S_1 is shown in figure 4 with the reverse current I as parameter. As in I it was found that the experimental spectra could be closely synthesized from a semi-empirical expression

$$S_1 = C + A/f + B(1+\omega^2\tau_1^2)^{-1} \qquad \ldots\ldots(1)$$

where $\omega = 2\pi f$; C is a uniform component, directly observable at high frequencies; A/f is an extensive component; $B(1+\omega^2\tau_1^2)^{-1}$ is localized in frequency in the region near $f=\tau_1/2\pi$ in so far as its experimental manifestation is concerned, where τ_1 is a relaxation time associated with the noise generation process.

The synthesis of the experimental spectrum for $I=70\,\mu\text{A}$ is shown in figure 5. Analysis of the family of spectra yields the data shown in the table. The dependence of the parameters C and A on I are shown in figures 6 and 7 respectively.

Reverse current I (10^{-6}A)	20	30	40	50	70	100	200
Reverse voltage V (v)	0·014	0·021	0·035	0·053	0·745	2·83	11·5
C (10^{-12}A)^2 sec.	6·4	9·6	12·8	16	22·4	32	64
A (10^{-9}A)^2				0·102	1·3	7·93	106
B (10^{-9}A)^2 sec.				2·58	9·33	69·6	1060
τ_1 (sec.)				0·1	0·08	0·2	0·26
$B/4\pi A\tau_1$				20·2	4·2	3·5	3
A/I^2				$4{\cdot}1 . 10^{-11}$	$2{\cdot}6 . 10^{-10}$	$7{\cdot}9 . 10^{-10}$	$2{\cdot}6 . 10^{-9}$

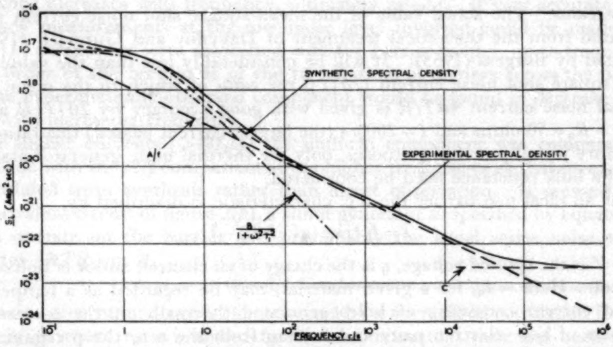


Figure 5. Synthesis of experimental excess noise density for $I=70\,\mu\text{A}$, $T=29°\text{C}$.

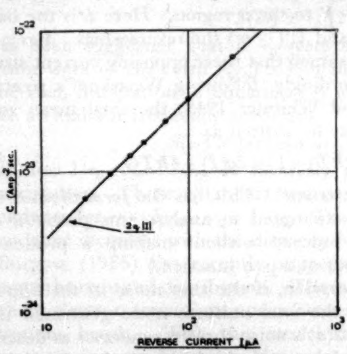


Figure 6. Dependence of uniform component C on reverse current.

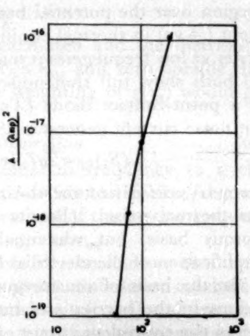


Figure 7. Dependence of A of $1/f$ component on reverse current.

§ 4. Discussion

(*a*) *Uniform Component*

The uniform component of excess noise is seen to be equal to $2q|I|$ for all currents, within experimental accuracy where $|I|$ is the *measured* current flowing in the diode. The *total* uniform component of mean square noise density $(S_1)_T$ is the power sum of this and the mean square thermal noise current density associated with the terminal conductance G_T

$$(S_1)_T = 2q|I| + 4\boldsymbol{k}TG_T. \qquad \ldots\ldots(2)$$

The near-ideal current range ($I < 60\,\mu\text{A}$) for which the equivalent circuit of figure 2(*a*) applies will be considered first.

To locate the fundamental noise generators the bulk resistance and barrier must be considered separately. The bulk resistance is expected to be the seat of thermal noise arising from the distribution of thermal velocities of electrons transversing the region and also of 'reduced shot noise' of the electron and hole streams. The actual value of the mean square shot noise current may be estimated from the theoretical treatment of Davydov and Gurevich (1943), as amended by Burgess (1955). It will be considerably less than the value of full mean square shot noise current ($2q|I|$) the ratio of which to the mean square thermal noise current $4\boldsymbol{k}T/R$ is given with good accuracy by $20|I|R$ at 29°c. For $R = R_0 = 50$ ohms and $I = 200\,\mu\text{A}$ (the largest current passed) this ratio is 0·2. Therefore for all practical purposes, only the thermal noise generator associated with the bulk resistance need be considered.

For an ideal p–n barrier the d.c. characteristic is described by

$$I = I_0\{\exp(qV/\boldsymbol{k}T) - 1\} \qquad \ldots\ldots(3)$$

where V is the applied voltage, q is the charge of an electron and $\boldsymbol{k}$ is Boltzmann's constant. Here $-I_0$, for a given material, may be regarded as a temperature-limited current consisting of holes generated thermally in the n-material as members of hole–electron pairs and flowing from the n to the p region. This is opposed by a hole current $I_0 \exp(qV/\boldsymbol{k}T)$ which flows from the hole-rich p region over the potential barrier $\phi - V$ to the n region. Here ϕ is the barrier height (+ve) in thermal equilibrium and V (−ve) the reverse bias. For a thin barrier at low frequencies it may be assumed that these opposing current streams will both show full shot noise. Accordingly, following Weisskopf's treatment for a point-contact diode (Torrey and Whitmer 1948) the total mean square shot noise current generator $(S_1)_{ST}$ may be written as

$$(S_1)_{ST} = 2qI_0\{\exp(qV/\boldsymbol{k}T) + 1\} = 2q(I) + 4\boldsymbol{k}TG_D \qquad \ldots\ldots(4)$$

where $G_D = \partial I/\partial V$ of the static characteristic. This has the *form* of shot noise plus thermal noise. (Petritz (1952) attempted to analyse this problem on a rigorous basis, but was unable to proceed without making a pronounced simplification of the electrical behaviour of a p–n junction.)

On the basis of the foregoing discussion, if the shot noise of the opposing streams in the barrier accounted for the total uniform noise generated there, then in the equivalent circuit of figure 2(*a*) a noise current generator as described by equation (4) would be included in shunt with the barrier at low frequencies. Furthermore a thermal noise voltage generator $4\boldsymbol{k}TR_0$ would operate in series with R_0. The above analysis is restricted to the frequency range for which the measured a.c. resistance of the diode is equal to the slope resistance of the static characteristic, i.e. for $\omega\tau_p < 1$ where τ_p is hole lifetime.[†] The lowest frequency at which comparison with experiment can be made in the present case, however, is 30 kc/s, because below this frequency the A/f component of noise manifests itself and the uniform component is not measured directly. At 30 kc/s the value of $\omega\tau_p$ is approximately 0·6 which is less than unity as required. The experimental results of equation (2) conform to this noise model at this frequency in that the measured terminal conductance G_T is very nearly equal to the barrier conductance G_D for all I and for ω less than $1/\tau_p$. In other words the contribution of the bulk resistance R_0 both to the measured admittance and to noise can be ignored for ω less than $1/\tau_p$.

For $\omega\tau_p$ greater than 1 the present experimental evidence is that the 'uniform' component remains so. It should be emphasized that this component is experimentally associated with the terminal admittance. As frequency increases R_0 makes an increasing contribution to the terminal impedance in that the barrier admittance increases with frequency, ultimately as $\omega^{1/2}$. If very accurate high-frequency measurements at the low noise level involved could be made, and if a $2q|I|$ component were associated with the barrier proper, up to frequencies of the order of the reciprocal of the transit time of carriers across the barrier, then the experimentally observed component would be found to decrease below $2q|I|$ with increasing frequency.

For higher currents ($>60\,\mu\text{A}$) the uniform component was comparable in magnitude with the A/f component at the higher frequencies so that its existence is postulated from synthesis rather than direct observation. It seems that in the equivalent circuit of figure 2(*b*), a shunt generator as specified by equation (4) should operate on the barrier and presumably the usual series noise voltage generator $4\boldsymbol{k}TR_0$ on R_0.

(*b*) *A/f Component*

A noise generation process characterized by a normalized correlation function $\exp(-\tau/\tau_0)$ where τ_0 is a well-defined mean relaxation time has been shown (Schottky 1926) to lead to a noise spectrum of the form

$$S(\omega) \propto \tau_0/(1+\omega^2\tau_0^2). \qquad \ldots\ldots(5)$$

It has been suggested, that if τ_0 were itself distributed and the distribution function were of the form $g(\tau_0) \propto 1/\tau_0$ for $\tau_{02} \leqslant \tau_0 \leqslant \tau_{01}$, and zero outside these limits, then integration of equation (5) with the weighting of $g(\tau_0)$ would give rise to a resultant frequency spectrum

$$S(\omega) \propto (\arctan \omega\tau_{01} - \arctan \omega\tau_{02})/\omega \qquad \ldots\ldots(6)$$

which for $1/\tau_{01} < \omega < 1/\tau_{02}$ is inversely proportional to frequency to a close approximation. There will in fact be departures of 3 dB from an inverse frequency dependence for $\omega_1 = 1/\tau_{01}$, $\omega_2 = 1/\tau_{02}$. The extensiveness of the inverse frequency dependence increases with the disparity between τ_{01} and τ_{02}.

Burgess (1955) has recently suggested a possible mechanism that might generate the required distribution. The model involves a modulation mechanism, in which electron transitions occur between bound levels (traps) and one of the continuous bands. With this model the long relaxation times required to generate the inverse frequency dependence at very low frequencies are readily explained, merely arising from transitions to high energies. The existence of a lower limiting relaxation time τ_{02} associated with the minimum energy transitions—those to the band edge—would seem to show that the high-frequency departure from inverse frequency dependence through the high-frequency form of equation (6)

[†] *Note added in proof.* van der Ziel (1955) has recently shown that equation (4) is applicable, independently of frequency restriction.

$$S(\omega) \propto \frac{1}{\omega} \operatorname{arc\,cot} \omega\tau_{02} \qquad \ldots\ldots(7)$$

might be encountered experimentally. This would require that τ_{02} was large enough for $1/\tau_{02}$ to lie within the frequency range in which non-uniform excess noise was detectable.

For the present device the so-called A/f component was detectable in some part of the frequency range for all currents from 50 μA upwards. The component was observed most extensively for $I = 200\,\mu$A, being measured directly over five decades of frequency from 10 c/s to 1 Mc/s. In this range the exponent of f was *indistinguishably different* from -1. It was not therefore possible to confirm the applicability of the model at frequencies for which equation (7) applies, i.e. for $\omega\tau_{02} \geqslant 1$, because presumably $1/\tau_{02}$ lay outside our upper frequency limit of 2 Mc/s.

It is believed that this is the most extensive frequency range (five decades) in which an experimental dependence of noise on f^{-1} has been directly observed (for a germanium device). In I, synthesis of the experimental spectra showed a basic component varying as f^{-1} over seven decades of frequency but this was not directly observed experimentally in that it was obscured by two 'bumps'. Synthesis in the present case also points to an f^{-1} dependence over seven decades of frequency.

The synthesis of the experimental spectrum at 100 μA on the basis of equation (1) did not give a very good fit below 300 c/s. This might well be due to the fact that above 300 c/s the experimental points refer to the 'basic level' of noise (switching was present at 100 μA) whereas below 300 c/s it was not possible to resolve the basic level and accordingly the observed mean noise level is plotted. The evidence points to the fact that the noise that was being switched was therefore of the A/f variety, in view of the fact that it was observed over such a wide range of frequencies.

The dependence of A on I shown in figure 7 is seen not to be of the form $A \propto I^2$, which would arise for any model (including that of Burgess) in which the current fluctuations arose from fluctuations in the resistance of the device, provided that any cross section normal to the current flow showed homogeneous resistance.

It is interesting to note, however, that if the d.c. characteristic of figure 3 is analysed in terms of an ideal diode of saturation current 60 μA, in shunt with a (surface) leakage path of ohmic resistance 68000 ohms, then in the region where leakage is apparent, i.e. for bias voltages greater than 0·5 volt, the value of A is proportional to the square of the excess current $(I-60)\,\mu$A. The generation of the A/f noise, for this particular diode at least, can therefore be interpreted as arising from fluctuations in the resistance of the leakage path across the junction rather than in the junction itself. In this case it is to be assumed, therefore, that the trapping centres, to and from which the electronic transitions responsible for the noise generation occur, reside in the surface leakage path across the junction.

(*c*) *Single Relaxation Time Component*

Only one component of this type was present in the noise spectra of the p–n junction diode, compared with two in those of the point contact diode of I. The relaxation time τ_1 in the present case gave rise to a 'bump' in the noise spectra in the frequency range between 0·1 and 100 c/s. In other words the relaxation time involved was of the order of one-tenth of a second.

There are certain similarities between the present component and those reported in I. The intensity B again increases with I but in this case somewhat less rapidly than A does. The relaxation time is again dependent on the current but here increases with I in contrast to the decrease of both the single relaxation times with increasing I for the point-contact diode.

The synthesis of the experimental spectrum for $I = 70\,\mu\text{A}$ in figure 5 again shows that the experimental bump is slightly broader than that associated with a single relaxation time. It is assumed therefore that this component arises from a noise generation process which is characterized by a relatively narrow spread of relaxation times centred on τ_1.

§ 5. CONCLUSION

The results presented pertain to a single p–n junction diode of the fused-alloy type. It was chosen primarily because it was a stable noise source over long periods for most bias currents ($I \simeq 100\,\mu\text{A}$ excluded). Unfortunately the d.c. characteristic and small-signal behaviour approached the ideal only for a limited low range of applied direct voltage; accordingly no specific conclusions may be drawn for p–n diodes as a class.

The synthesis of the experimental spectra from the same basic spectral components as were required for a point-contact diode (see I) is significant, as is the slope, indistinguishable from -1, of the extensive component. The existence of a component of this type, sometimes only loosely called an inverse frequency component because the exponent of f has been only nearly equal to -1, has been widely reported for germanium devices, although sometimes the frequency range of measurement has been only two or three decades.

It has been established that the inverse frequency component of noise is probably generated in the surface leakage component of current, rather than in that crossing the p–n junction.

The closeness of the fit of the experimental points to the line $2q|I|$ for the uniform component might be somewhat fortuitous for the low currents (i.e. 20 and $30\,\mu\text{A}$). Further work at low currents and high frequencies is undoubtedly required. This would be more easily done with a diode having a higher zero-voltage resistance than the present one so that the ratio of the excess to thermal noise voltage $20\,IR$ (at 29°c) was increased. Its stability as a noise source is presupposed.

ACKNOWLEDGMENTS

The author wishes to acknowledge the assistance of Mr. H. C. Bevan with the experimental work. The work described above was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

REFERENCES

BURGESS, R. E., 1941, *Proc. Phys. Soc.*, **53,** 293; 1955, *Brit. J. Appl. Phys.*, **6,** 185.

DAVYDOV, B., and GUREVICH, B., 1943, *J. Phys.*, *U.S.S.R.*, **7,** 138.

HYDE, F. J., 1953, *Proc. Phys. Soc.* B, **66,** 1017.

PETRITZ, R. L., 1952, *Proc. Inst. Radio Engrs.*, **40,** 1440.

SCHOTTKY, W., 1926, *Phys. Rev.*, **28,** 74.

SHOCKLEY, W., 1949, *Bell Syst. Tech. J.*, **28,** 435.

TORREY, H. C., and WHITMER, A., 1948, *Crystal Rectifiers* (Radiation Laboratory Series of Massachusetts Institute of Technology, **15**), (New York and London: McGraw-Hill).

VAN DER ZIEL, A., *Proc. Inst. Radio Engrs, N.Y.*, **43,** 1639.

APPENDIX I

In figure 8 the diode and feed circuit are represented by an excess noise voltage generator *e* (the 'signal' in these measurements) operating in series with resistance *r*, which is shunted by capacitance *c*; *s* is the equivalent noise input resistance of the receiver. c_1, c', L and R comprise the coupling network (R is the equivalent shunt resistance arising from the resistance of the coil). The noise figure F is a measure of the deterioration of the signal to noise ratio at the terminals of the source when succeeded by a receiver (say) which itself generates noise. It may be conveniently defined as the ratio of total noise output from the receiver to that noise output due to the source alone.

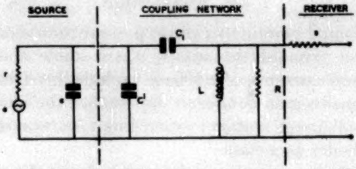


Figure 8. The coupling of source to receiver to give low noise figure.

For the present case, in the absence of the coupling network this is

$$F = 1 + s(1 + \omega^2 c^2 r^2)/r;$$

with the coupling network inserted the noise figure can be shown to have an optimum value F_{opt} when $r' = R\{s/(R+s)\}^{1/2}$ where $r' = r\{(c_1 + c_2)/c_1\}^2 + (\omega^2 r c_1^{\,2})^{-1}$ and $c_2 = c + c'$. This is

$$F_{\text{opt}} = 1 + \frac{2s}{R}\left[1 + \left(\frac{R+s}{R}\right)^{1/2}\right]$$

which can represent a considerable improvement, particularly if R/s can be made large. For example if $s = 5000$ ohms, $r = 100$ ohms, $c = 1000$ pF, $f = 10^6$ c/s, $R = 20\,000$ ohms the noise figures with and without the optimized coupling network are approximately 2 and 70. The coupling network used here is particularly useful for capacitative sources. The general principle of improving signal to noise ratio by adjusting source–amplifier coupling was first illustrated by Burgess (1941).

APPENDIX II

Small-Signal Equivalent Circuit

Analysis of the terminal admittance components, together with the d.c. characteristic showed that, for frequencies up to 2 Mc/s, the behaviour of the diode for reverse bias could be explained on the basis of the one-dimensional diffusion model for an abrupt p–n transition (Shockley 1949) with an associated depletion region, and a series resistance R_0 arising in the bulk (n-type) material of the device. The diode was far from ideal. There was no saturation of the reverse current. For low direct voltages ($<6\times10^{-2}$ volt), however, a good fit to the ideal form of the d.c. characteristic of equation (3) was obtained if I_0 was taken as $60\,\mu\mathrm{A}$. The non-saturation of the reverse current was considered to be partly due to a flow of current from the depletion region, which by its nature would be the seat of the high rate of generation of thermally excited hole–electron pairs. For uniform material the current arising in the depletion region would be proportional to the volume (width) of the region and would increase with applied reverse voltage. The shunt conductance associated with this current flow is designated G_{S}. The depletion region would also give rise to a Schottky space-charge capacitance C_{S} virtually equal to the measured barrier capacitance for large reverse voltages. The ratio of capacitance to conductance of such a region should be constant for uniform material. For currents $I>60\,\mu\mathrm{A}$ for which this constancy might have been expected it was found that the product of the resistance and capacitance of the barrier decreased with increasing voltage (diffusion capacitance and conductance are negligible in this range). A plot of $1/C_{\mathrm{S}}^2$ against V, however, did give the straight line to be expected for a Schottky type barrier capacitance. To give better correspondence with experimental results for $I>60\,\mu\mathrm{A}$ it would be necessary to invoke either non-uniform material in the depletion region or a leakage a.c. resistance of about 200 000 ohms across the barrier. This would, however, conflict with the value of the d.c. leakage resistance (68 000 ohms) invoked in the analysis of the d.c. characteristic.

On the basis of this model the small-signal equivalent circuit is shown in figures 2(*a*) and 2(*b*). Here R_0 is the bulk resistance, $R_{\mathrm{D}}=1/G_{\mathrm{D}}$ the diffusion resistance, C_{D} the diffusion capacity, $R_{\mathrm{S}}=1/G_{\mathrm{S}}$ the depletion region resistance, C_{S} the depletion region capacitance, $R_{\mathrm{L}}=1/G_{\mathrm{L}}$ the possible leakage resistance. The parameters giving the best fit to the experimental admittance curves are hole lifetime $\tau_{\mathrm{p}}=3{\cdot}1\,\mu\mathrm{sec}$, $R_0=50$ ohms, $R_{\mathrm{D}0}=1/G_{\mathrm{D}0}$ (diffusion resistance for zero applied voltage and $\omega\tau_{\mathrm{p}}\ll1)=450$ ohms. R_{D} and C_{D} for any frequency then follow (Shockley 1949) from

$$G_{\mathrm{D}}+i\omega C_{\mathrm{D}}=(1+i\omega\tau_{\mathrm{p}})^{1/2}G_{\mathrm{D}0}\exp(q\,|V|/\boldsymbol{k}T).$$

RESEARCH NOTES

Excess Noise Spectra in Germanium

By F. J. HYDE
Radio Research Station, Slough, Bucks.

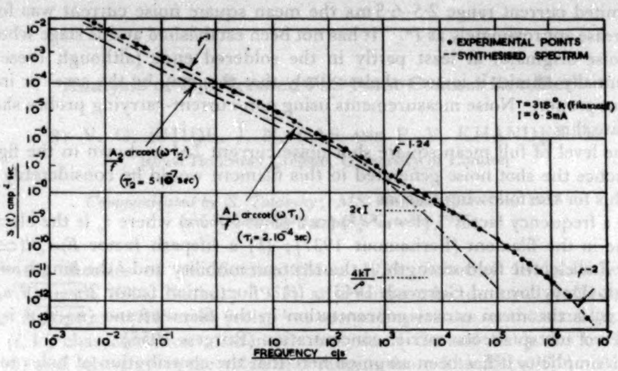
Communicated by R. L. Smith-Rose; MS. received 16*th August* 1955
and in final form 1*st November* 1955

It has been found experimentally by several investigators that there is generated in germanium devices a component of electrical noise (in excess of the thermal noise associated with the real part of the admittance) whose spectral density $S(f)$ decreases with increasing frequency. By synthesis of the experimental noise spectra of a point-contact germanium diode it has been shown (Hyde 1953) that an extensive component of this type can vary as f^{-1} over as many as seven decades of frequency (10^{-1}–10^{6} c/s). In more recent work by Hyde (to be published) a dependence of noise density on frequency with an exponent indistinguishably different from -1 has been directly observed over five decades of frequency (10–10^{6} c/s) for a p–n junction diode. Montgomery's (1952) work on transistors, diodes and filaments showed an approximate dependence on f^{-1} although the exponent was usually slightly greater than unity: his frequency range was, however, somewhat limited.

In postulating a theory of noise of which the spectral density varies as f^{-m} over a wide range of frequencies, consideration must be given as to the extent to which a value of m different from unity is fundamental to the basic hypothetical mechanism. Bess (1953) has proposed a theory which is based on diffusion of impurity atoms in the surface layer (thickness δ, diffusion constant D) and the annihilation of the vacancies (which act as traps) produced by the removal of the diffusing atoms from their equilibrium sites. By considering vacancies to decay either through capturing their original atoms or others (σ_c = capture cross section of a vacancy, $\bar{v}$ = average velocity of diffusing atoms) he was able to arrive at a spectrum which can vary as f^{-m} where $m=2-n\simeq 1$, $n=\sigma_c\bar{v}/4\pi\delta D$. As two different processes are involved here it would be very unlikely for n to be equal to unity for all specimens; indeed the theory is formulated to cover frequency dependence where the exponent can differ slightly from unity over several decades of frequency. A unique exponent is, however, implied for a particular specimen.

Burgess (1955) has envisaged a modulating mechanism involving hole and electron transitions from recombination centres to a range of energies in the valence or conduction bands. The probability of a transition to a level differing by ϵ electron volts from that of a centre is proportional to $\exp(-|\epsilon|/\boldsymbol{k}T)$ and the relaxation time τ_0 of the transition is proportional to $\exp(|\epsilon|/\boldsymbol{k}T)$, so that the probability of any τ_0 is proportional to $1/\tau_0$. (A minimum relaxation time τ_1 is involved; this arises from transitions from a recombination centre to the band edge.) Such a distribution of relaxation gives rise to a spectrum of the form $f^{-1}\operatorname{arc\,cot}(\omega\tau_1)$ where $\omega=2\pi f$. For frequencies much less than $1/\tau_1$ the spectrum will vary as f^{-1} while for frequencies much greater than $1/\tau_1$ the spectrum will decrease as f^{-2}.

The diagram shows the dependence of the spectral density of the mean square noise current generator of an aged two-terminal n-type germanium filament (dimensions 0·84 cm × 0·55 mm × 0·65 mm, current $I = 6{\cdot}5$ mA, voltage $V = 16$ v, a.c. resistance at 1 kc/s $\rho = 2470$ ohms, zero-frequency resistance 1200 ohms, thermal time constant 2·3 seconds, minority carrier lifetime measured in the filament 1·7 μsec, uniform resistivity 10·5 ohm cm at 45°c. Corresponding to electron and hole concentrations of $1{\cdot}4 \times 10^{14}$ and $2{\cdot}3 \times 10^{13}$ cm^{-3} respectively.



Synthesis of excess noise spectrum of germanium filament.

Note. In ordinate, read μamp for amp.

The disparity between the a.c. resistance at 1 kc/s, say, and the zero-frequency resistance arises because of the temperature rise produced in the filament through internal power dissipation in it, which gives rise to a non-linear direct current–voltage characteristic. In consequence the impedance of the filament becomes frequency dependent at frequencies of the order of the reciprocal of the thermal time constant. The determination of the relevant impedance of the filament at any frequency is described in a separate paper (Hyde, to be published). It is seen that within experimental accuracy the negative exponent of frequency changes from one at low frequencies to two at high frequencies and in the transition range between these values is sensibly constant over three decades of frequency ($50 - 5 \times 10^4$ c/s) having an apparent value of 1·24. It may be noted that at 4 Mc/s the excess noise measured was only 5% of the thermal noise $4\boldsymbol{k}T/\rho$. The synthesis of this spectrum is shown from two f^{-1} arc cot $(\omega\tau)$ spectra with limiting relaxation times τ_1, τ_2 of 2×10^{-4} and 5×10^{-7} second. These times and the intensities (A_1, A_2) have been chosen to give the best fit. It is interesting to note that by superposing only two basic spectra of this type, such as are postulated to arise from two recombination centre levels, it is possible to generate an exponent which is different from unity (1·24 in this case) and practically constant over three decades of frequency. (An exponent less than unity cannot arise from superposing spectra of the form f^{-1} arc cot $(\omega\tau)$, but in the author's experience exponents less than unity have been satisfactorily explained in terms of a component of spectral density of the form $B/\{1 + (\omega\tau_B)^2\}$, where τ_B is a discrete relaxation time of the noise generation process.)

That the high frequency departure of the 'inverse frequency' noise from an f^{-1} to an f^{-2} dependence has been measured is in some measure due to the absence of 'bump' components arising from discrete relaxation times at high frequencies. Also it is perhaps fortunate that the limiting relaxation times τ_1 and τ_2 have reciprocals inside the frequency range in which it is possible to measure excess noise.

Spectra of similar shape were obtained for currents of 2·5 and 4·5 mA. For the limited current range 2·5–6·5 mA the mean square noise current was found to increase approximately as I^3. It has not been established at this stage whether the noise originates at least partly in the soldered ends (although these are substantially ohmic, it is nevertheless likely that this may be the case) or in the filament proper. Noise measurements using non-current-carrying probes should elucidate this.

The level of full mean square shot noise current $2eI$ is shown in the figure. In practice the shot noise generated in this filament would be considerably less than this for the following reasons:

(*a*) a frequency factor $1/(1+\omega^2\tau_e^2)$ ($\simeq 0{\cdot}45$ at 10^5 c/s) where τ_e is the electron lifetime in the filament (Bernamont 1937); (*b*) a lifepath factor $E\mu\tau_e/l(\simeq 0{\cdot}1)$ where E is electric field strength, μ the electron mobility and l the length of the filament (Davydov and Gurevich 1943); (*c*) a fluctuation factor $2\overline{(n-n_0)^2}/n_0<1$ where n_0 is the mean carrier concentration in the filament and $\overline{(n-n_0)^2}$ is the variance of instantaneous carrier concentration (Burgess 1955).

For simplicity it has been assumed here that the contribution of holes to the conductivity can be neglected. The filament temperature with a current of 6·5 mA flowing through it was 45°C so that this assumption does amount to an approximation. The order of magnitude of the reduction factors, however, is not affected.

It seems likely therefore that shot noise does not make a significant contribution to the observed spectra.

It may be remarked that the generation of a sensibly constant value of m greater than 1 and less than 2 over more than three decades of frequency could readily arise if more than two recombination centre levels were present, and the individual intensities and limiting relaxation times of the $f^{-1}\operatorname{arc\,cot}(\omega\tau)$ spectra were compatible.

In experiments in p–n junction photocells reported by Shive (1952) it was noted that carriers produced by illumination showed only shot noise, the inference being that the 'inverse-frequency' noise which was also present was essentially bound up with the thermal generation of carriers and not with a modulating process which could affect all carriers irrespective of their source. This, together with the present experimental evidence, seems to favour the Burgess hypothesis as the source of the inverse frequency noise if this can be assumed the same for all germanium devices.

ACKNOWLEDGMENTS

The author is indebted to his colleagues Messrs. H. C. Bevan and R. W. Smith for assistance in the experimental investigation. The work described above was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

References

Bernamont, J., 1937, *Proc. Phys. Soc.*, **49**, E 138.
Bess, L., 1953, *Phys. Rev.*, **91**, 1569.
Burgess, R. E., 1955, *Brit. J. Appl. Phys.*, **6**, 185.
Davydov, B., and Gurevich, B., 1943, *J. Phys. U.S.S.R.*, **7**, 138.
Hyde, F. J., 1953, *Proc. Phys. Soc.* B, **66**, 1017.
Montgomery, H. C., 1952, *Bell Syst. Tech. J.*, **31**, 950.
Shive, J. N., 1952, *Proc. Inst. Radio Engrs, N.Y.*, **40**, 1410.

Analysis of H_2S Treated PbS Point Contact Rectifiers

By V. G. BHIDE, J. N. DAS and P. V. KHANDEKAR
Royal Holloway College, University of London

*Communicated by S. Tolansky; MS. received 12th April 1955
and in amended form 6th October 1955*

The theories of rectification (Mott 1939, Schottky 1941, Bethe 1942, Courant 1946, Banbury 1953) at a semiconductor–metal contact expect a slope $\alpha = e/\boldsymbol{k}T$ or 40 (volt)$^{-1}$ at room temperature in the semi-log plot of the (i, V) characteristics. Extensive analysis of Ge and Si diodes by Yearian (1950) and others gives the value of α ranging from 5 to 35 (volt)$^{-1}$. Although much work has been done on the diode and triode action in PbS (Henisch and Granville 1950, p. 87, Banbury and Henisch 1950, Gebbie, Banbury and Hogarth 1950, Gibson 1952, Reimann and Sullivan 1952) no attempt has so far been made to compare the observed characteristics with the theoretical model. In the present studies, natural and H_2S treated PbS rectifiers were subjected to a detailed analysis with a view to test the applicability of various theories and also to determine the effect of p-type layer on n-type material.

Natural and synthetic crystals which showed n type conductivity were heated to about 500°c in an atmosphere of flowing H_2S for about half-an-hour. The crystals were then quenched. These treated samples showed p-type conductivity. Similar conversion of n-type material to p-type and vice versa was first reported by Eisenmann (1940) and Hintenberger (1942). These workers were concerned with the bulk conversion and hence subjected their samples to severe sulphurization. We treated the crystals for different durations thereby leaving a p–n junction in the body of the crystal, the depth of which was determined by etching the crystal and testing the sign of rectification, thermal e.m.f. and photovoltaic effect (Granville and Hogarth 1951) at different depths. The depth of etch pits was determined by light profile microscopy (Tolansky 1952). In this technique, the field of view of the microscope is crossed by a dark profile line which is the image of an opaque line placed at the field iris and projected on the surface by an off-centre illumination. The profile gives the line contour of the specimen where changes in level are registered as lateral shifts of the profile as seen in the figure (Plate) in which the depth of etch pit corresponds to $0{\cdot}55\,\mu$. The depth of p–n junction was found to be a linear function of time tending gradually towards saturation for a given rate of flow and crystal temperature.

Crystal	Whisker	$T(^\circ\text{K})$	Slope	$S=\alpha(\boldsymbol{k}T/e)$	R_s (ohms)	ϕ (v)	$(i_0)_f$ (μA)	$\beta e/\boldsymbol{k}T$	β	$1-\beta$	$(i_0)_b$ (μA)
A_1P	Ni	306·4	35	0·91	161	0·18	2·1	1·8	0·045	0·95	—
A_1P	Ni	311·0	32·8	0·85	125	0·25	0·11	2·54	0·066	0·93	0·2
A_1P	Phos. Broz.	311·0	35·7	0·93	352	0·18	0·29	0·98	0·025	0·98	0·35
A_1P	W	311·0	37·0	0·96	147	0·19	0·31	0·95	0·025	0·98	0·45
A_1P	W	310·9	38·4	0·99	116	0·20	0·23	3·07	0·080	0·92	0·2
A_2P	Ni	313·8	32·0	0·84	549	0·22	0·50	2·88	0·075	0·93	0·50
A_2P	Ni	313·8	32·5	0·85	440	0·26	0·05	4·19	0·110	0·89	0·05
A_2P	W	313·7	37·7	0·98	448	0·19	0·11	1·99	0·052	0·95	0·20
A_2P	Phos. Broz.	313·4	38·9	1·02	1300	0·15	0·15	2·53	0·066	0·93	0·20
A_3P	Ni	313·2	41·1	1·07	175	0·12	0·30	0·57	0·015	0·99	1·0
A_4P	W	312·6	38·7	1·01	128	0·25	0·01	—	—	—	0·15
A_4P	W	313·0	40·0	1·04	680	0·16	0·16	—	—	—	0·30
A_4P	Phos. Broz.	313·8	39·7	1·04	480	0·22	0·06	1·61	0·042	0·96	0·10
I(N)	Ni	306·0	9·9	0·25	53	0·52	12·3	3·85	0·098	0·90	—
II(N)	Ni	306·0	3·6	0·09	43	0·75	77·3	4·50	0·12	0·88	—

P=p type; N=n type

Points which made good rectifying contacts were investigated. Whiskers of different materials made from wires 1 to 3 mils and pointed by electrolytic etching were engaged to the crystal under optimum pressure (Das and Bhide 1954). Results of the investigation of the static characteristics of n-type and treated samples are recorded in the table.

The conductivity of natural n-type specimens was found to be of the order 400 $\text{ohm}^{-1}\,\text{cm}^{-1}$ which suggests that the crystals are sufficiently impure to inhibit carrier injection. This is borne out by the large reverse currents and poor diode characteristics. In the absence of such carrier injection, the spreading resistance R_s and the barrier height were determined by extrapolating the linear portion of (i, V_a) curves in the easy flow direction. The R_s thus determined and its corresponding values obtained following Gibson (1952) are in good agreement. It is interesting to note that ϕ decreases on partial sulphurization. It, however, increases after complete conversion. The slope α of the semi-log plot $[i, (V_a - iR_s)]$ for n-type material was about 10, a quarter of the theoretically expected value. Unlike the n-type specimen, and also Ge and Si diodes, the converted samples yielded the theoretical value of α with all the whiskers studied. The slope of the linear portion of the reverse semi-log curve $\beta e/\boldsymbol{k}T$ (Courant 1946) agrees fairly well with its value calculated from the forward characteristics [column S in the table]. It is also of some interest to note that the extrapolated zero voltage currents in the forward direction $(i_0)_f$ and the saturating currents in the reverse direction $(i_0)_b$ are of the same order. The treated crystals could stand higher breakdown voltage and exhibited more stable characteristics.

Yearian has explained the divergence between the observed and the theoretical values of α by suggesting that the actual contact covers a number of points with different values of ϕ. This postulate does not lead to unique determination of ϕ and its distribution within the contact area. The data presented here indicates that if $\phi < 0{\cdot}25$ volts, the slope of $(\ln i, V)$ curves almost equals $e/\boldsymbol{k}T$. Conversely if ϕ exceeds this value as for n-type material, the slope is much less than $e/\boldsymbol{k}T$. In this connection it is of interest to note that the slope of $(\ln i, V)$ curves for Ge and Si diodes rapidly changes from theoretically expected values to much lower values after 0·2 to 0·3 volts (Torrey and Whitmer 1948).

The observed correspondence of the treated samples with the theoretical model may alternately be due to the existence of p–n junction in the body of the crystal.†

Our thanks are due to Dr. V. N. Thatte for his help. One of us (V.G.B.) wishes to thank Professor S. Tolansky, in whose laboratory at Royal Holloway College part of the work was done, for his kind interest and encouragement and also the Ministry of Education, the Government of India, State of Madhya Pradesh and Nagpur University for award of the scholarship.

REFERENCES

BANBURY, P. C., 1953, *Proc. Phys. Soc.* B, **66,** 833.
BANBURY, P. C., and HENISCH, H. K., 1950, *Proc. Phys. Soc.* B, **63,** 540.
BETHE, H. A., 1942, *N.D.R.C. Division* 14 *Report* 43 (Massachusetts Institute of Technology).
COURANT, E. D., 1946, *Phys. Rev.*, **69,** 684.

---

† This is suggested by the fact that in these crystal rectifiers the condition of easy flow is obtained when the crystal is negative with respect to the whisker, indicating that the seat of rectification is at the p–n junction.

DAS, J. N., and BHIDE, V. G., 1954, *Curr. Sci.*, **23**, 185.
EISENMANN, L., 1940, *Ann. Phys., Lpz.*, **38**, 121.
GEBBIE, H. A., BANBURY, P. C., and HOGARTH, C. A., 1950, *Proc. Phys. Soc.* B, **63**, 371.
GIBSON, A. F., 1952, *Proc. Phys. Soc.* B, **65**, 196.
GRANVILLE, J. W., and HOGARTH, C. A., 1951, *Proc. Phys. Soc.* B, **64**, 488.
HINTENBERGER, H., 1942, *Z. Phys.*, **119**, 1.
HENISCH, H. K., and GRANVILLE, J. W., 1950, *Semiconducting Materials*, Ed. H. K. Henisch (London: Butterworths Scientific Publications), p. 87.
MOTT, N. F., 1939, *Proc. Roy. Soc.* A, **117**, 27.
REIMANN, A. L., and SULLIVAN, J. V., 1952, *Proc. Phys. Soc.* B, **65**, 480.
SCHOTTKY, W., 1941, *Z. Phys.*, **118**, 539.
TOLANSKY, S., 1952, *Z. Elektrochem*, p. 263.
TORREY, H. C., and WHITMER, C. A., 1948, *Crystal Rectifiers* (New York: McGraw Hill).
YEARIAN, H. J., 1950, *J. Appl. Phys.*, **21**, 214, 283.

The Diffusion of Chromium in Titanium studied by an Autoradiographic Method

BY A. J. MORTLOCK AND D. H. TOMLIN
Physics Research Laboratory, University of Reading

Communicated by R. W. Ditchburn; MS. received 26th September 1955

AUTORADIOGRAPHIC methods have been used by several workers to study diffusion in solids, and measurements on the self diffusion of gold (Gatos *et al.* 1952, 1954) and silver (Krueger and Nersh 1955) have given results in agreement with those obtained by more conventional methods.

An autoradiographic method has been used by the present authors for studying the diffusion of chromium in titanium. The purpose of this note is to show that the method allows several diffusion experiments to be carried out on a single diffusion sandwich.

The sandwiches were formed by evaporating Cr containing the pile-produced radioactive isotope ^{51}Cr (26 day half-life) on to one finely ground end face of each of two small cylindrical specimens of Ti (commercial grade) or Ti + 18% Cr, and clamping the cylinders together with the activated ends in close contact by means of a small titanium bolt passing through an axial hole. A longitudinal flat surface had previously been ground on each cylinder, and in assembling them, these longitudinal surfaces were aligned. They were afterwards ground a little more to produce a surface cutting the radioactive interface normally and free from stray chromium deposits.

An autoradiograph of this surface was prepared on x-ray film to show the sharpness with which the initial layer of chromium could be detected. The soft electron radiation from ^{51}Cr was found to provide good autoradiographic resolution under these experimental conditions.

The sandwich was then annealed in a temperature controlled vacuum furnace for a known time after which the longitudinal surface was ground down by a further 1 to 2 mm to remove surface diffusion effects, and a second autoradiograph prepared, showing a diffused zone of chromium in place of the initial sharply defined interfacial layer. The variation of the photographic density of the image in the direction normal to the interface was measured by means of a high resolution scanning microphotometer. The corresponding variation of chromium concentration was deduced from the previously established linearity of the density–exposure relation of the x-ray film.

Further diffusion runs were then carried out in a similar way, autoradiographs being prepared after each, and the chromium concentration distribution determined.

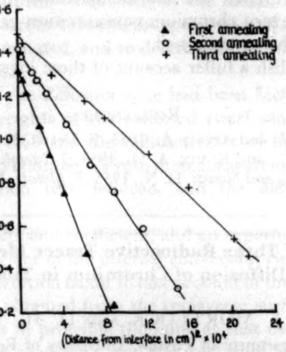
If $c(x)$ denotes the chromium concentration as a function of the distance x from the interface after annealing for time t at a given temperature, then Fick's law gives

$$\ln c(x) = -x^2/4Dt + \text{const.}$$

provided that the diffusion coefficient D is a constant. It is easily shown that after successive diffusion runs of duration $t_1, t_2 \ldots$ etc.

$$\ln c(x) = -x^2/4(D_1t_1 + D_2t_2 + \ldots) + \text{const.}$$

where $D_1, D_2 \ldots$ refer to the corresponding annealing temperatures. The gradients of the plots of $\ln c$ against x^2 made after each run in a multiple experiment of the kind described give the values of D_1t_1, $D_1t_1 + D_2t_2$, etc.



Penetration of diffused Cr in Ti+18%Cr alloy after each of three successive annealing treatments. Photographic density is proportional to Cr concentration.

The figure shows typical plots for three successive runs with a Ti + 18% Cr sandwich. The measured concentration distributions were symmetrical about the interface to within experimental error, and the points plotted are means of corresponding readings on both sides. Results derived from this and from a separate double experiment on a similar sandwich are given in the table. They show no significant variation of diffusion coefficient among four runs at 1000°c, and the result for 1047°c is in agreement with other data so far obtained over a limited temperature range.

Diffusion Sandwich No.	1	1	1	2	2
Diffusion Run No.	1	2	3	1	2
Temperature (°c)	1047	1000	1000	1000	1000
Time (hours)	6	10	10	10	10
Diffusion coefficient (cm² sec⁻¹ × 10⁹)	3·9	1·8	2·2	2·0	2·0

It is possible that a greater number of runs could be carried out on one specimen though a limit is set by the progressive loss in specific activity due to diffusion and radioactive decay, and by increasing experimental error.

Nevertheless, considerable advantages derive from the possibility of carrying out a limited number of successive experiments, mainly in establishing the validity of a measured diffusion coefficient, and in studying its dependence on factors such as grain size and mean solute concentration, which change during heat treatment. The present results indicate that in the alloy used, at a temperature of 1000°c, the successive heat treatments did not significantly affect the measured diffusion coefficient. The mean grain diameter was initially 0·07 mm and increased by a factor of about 3 during the first two runs, but only by a much smaller factor in the third run. The concentration of chromium was 100% in the initial interfacial layer, but since this layer was only about 0·1 micron thick, was reduced to a value very close to the uniform 18% level within a comparatively short time at the beginning of the first anneal. No progressive contamination of the alloy by oxygen or nitrogen was indicated by hardness measurements carried out before and after each annealing treatment.

The experiments carried out on the diffusion of chromium in titanium cylinders at effectively zero chromium concentration gave less consistent results, reasons for which are being examined.

It is hoped to publish a fuller account of these investigations at a later date.

REFERENCES

GATOS, H. C., and AZZAM, A., 1952, *J. Metals, N.Y.*, **4,** 407.
GATOS, H. C., and KURTZ, A. D., 1954, *J. Metals, N.Y.*, **6,** 616.
KRUEGER, H., and NERSH, H. N., 1955, *J. Metals, N.Y.*, **7,** 125.

A Comparison of Three Radioactive Tracer Methods for Studying the Diffusion of Chromium in Titanium

BY A. J. MORTLOCK AND D. H. TOMLIN
Department of Physics, University of Reading

Communicated by R. W. Ditchburn; *MS. received 3rd November* 1955

§ 1. INTRODUCTION

AUTORADIOGRAPHIC methods have been used by several workers to measure diffusion coefficients in metallic systems. In particular, the present authors (Mortlock and Tomlin 1956), recently described its application to the titanium chromium system, pointing out some of its advantages over other tracer methods.

The only comparisons of the autoradiographic and other methods which have so far been reported are between quite independent investigations of the self-diffusion in gold (Gatos and Kurtz 1954) and silver (Krueger and Nersh 1955). As a comparison was also considered necessary in the present work, an experiment was carried out in which the different methods of analysis were applied to a single diffusion specimen. In this way a true comparison of methods was achieved, and all experimental errors other than those involved in analysing the diffusion zone were eliminated.

§ 2. EXPERIMENTAL METHODS

As previously reported, the diffusion specimen was in the form of a sandwich, in which a thin interfacial layer of spectrographically pure chromium containing the radioactive isotope ^{51}Cr was located between the ends of two well-annealed titanium cylinders of commercial purity.

A diffusion heat treatment at 1055° ± 4°c was carried out *in vacuo* for 24 hours, after which the specimen was ground so as to produce a longitudinal flat surface perpendicular to the interface. An autoradiograph of this surface was prepared, and the distribution of chromium concentration determined by microdensitometry; from this a diffusion coefficient was calculated.

The specimen was next mounted in a lathe and serial layers parallel to the interface and about 0·1 mm thick were machined off, a known proportion of the swarf representing each layer being collected in a suction trap. After the removal of each layer, the length of the specimen was measured with a micrometer.

Before remounting in the lathe for removal of the next layer, the radiation from the newly exposed transverse face was measured with a thin end-windowed Geiger counter. This radiation consisted of soft internal conversion electrons having a high self-absorption coefficient in the metal, and penetrating γ-rays. By placing a suitable absorber between the specimen and the counter, the electron component could be eliminated, and so determined by subtracting the remaining γ-ray count from the total.

When the whole of the diffusion zone had been sectioned and measured in this way, the known weights of the collected swarf samples were dissolved in 12 ml. of concentrated HCl and the radioactivities of the solutions determined with a liquid type Geiger counter. From these measurements a second chromium concentration distribution was deduced, and the diffusion coefficient again calculated.

This value of the diffusion coefficient, and an experimentally estimated value of the absorption coefficient of the internal conversion electrons in titanium were used in calculating a correction factor to take account of the effect of self absorption on the electron counts obtained from the transverse surfaces.

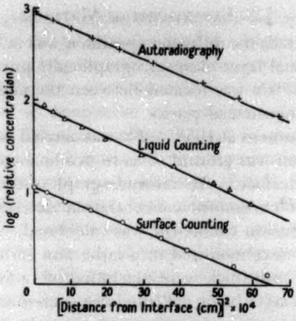
The calculation was in principle the same as that used by Gatos and Kurtz (1954). An absorption coefficient was deduced from absorption data obtained with a thin source of radioactive chromium and absorbers of thin card, with the assumption that thickness of card and of titanium metal are equivalent when expressed in mg cm^{-2}.

The surface counting data therefore provided a third method for determining the chromium concentration distribution in the diffusion zone, and hence the diffusion coefficient.

§ 3. RESULTS

In each case the diffusion coefficient was determined from the slope of a plot of log (relative concentration) against (distance from interface)^2. The three plots are shown in the figure, the points being means of both sides of the corresponding distributions. They are linear to within experimental error, and lead to values of diffusion coefficient as follows:

Experimental Method	Diffusion Coefficient
Autoradiographic	$7{\cdot}4\times10^{-9}$ cm^2 sec^{-1}
Counting dissolved layers	$7{\cdot}0\times10^{-9}$ cm^2 sec^{-1}
Counting transverse surfaces	$7{\cdot}8\times10^{-9}$ cm^2 sec^{-1}



Penetration of diffused Cr in Ti determined by autoradiography, liquid counting, and surface counting.

§ 4. CONCLUSION

It is estimated that the experimental error in the values of the diffusion coefficient determined from each of the three methods is between 5% and 10%; the values therefore agree satisfactorily. It should be pointed out that the autoradiographic and transverse surface counting methods used in this work were made possible because of the presence of a soft electron component in the ^{51}Cr radiations. When, as in this case, the type of radiation from a particular isotope is such that the necessary resolution is obtainable, the autoradiographic method of studying diffusion is reliable and convenient.

REFERENCES

GATOS, H. C., and KURTZ, A. D., 1954, *J. Metals*, **6**, 616.
KRUEGER, H., and NERSH, H. N., 1955, *J. Metals*, **7**, 125.
MORTLOCK, A. J., and TOMLIN, D. H., 1956, *Proc. Phys. Soc.* B, **69**, 248.

Absence of Radio Frequency Resonance Absorption in Triethylamine

BY J. S. DRYDEN AND R. J. MEAKINS
Division of Electrotechnology, National Standards Laboratory, Australia

MS. received 2nd September 1955, *and in amended form* 18*th October* 1955

DAVIES AND PRICE (1954) and Davies (1954) have claimed recently that the dielectric properties of solutions of triethylamine in nujol (medicinal paraffin) indicate the presence of a resonance absorption centred about a frequency of 80 Mc/s. It was suggested that this absorption is similar to the well-known absorption in gaseous ammonia at a frequency of 24 Gc/s (kMc/s). The question of whether this is a resonance effect or not is of considerable importance, particularly since there is, as yet, no accepted resonance absorption in a solid or liquid at radio frequencies. The present authors therefore set out to repeat the measurements of Davies and Price and to carry out a more complete investigation into this absorption in triethylamine.

The triethylamine, b.p. 89·4°c, was obtained from commercial material by fractional distillation through a 30 cm column packed with stainless steel gauze cylinders.

Measurements between 30 Mc/s and 120 Mc/s were made by a susceptance variation method in which the liquid was placed between two dished electrodes similar to those described by Hartshorn and Ward (1936). Between 100 Mc/s and 300 Mc/s a re-entrant cavity described by Hollway and Cassidy (1952) was used, the *Q* being again measured by susceptance variation. In this apparatus the liquid was held between parallel plate electrodes in a shallow fused silica cup. Standing wave methods (Roberts and von Hippel 1946) were used at frequencies of 1·1, 3 and 8·6 Gc/s, precautions being taken to minimize errors due to the meniscus. At each frequency the solution was compared with the solvent and the incremental loss factor determined.

The results obtained in the present work for a 0·48 M solution of triethylamine in nujol are shown in figure 1 together with the curve published by Davies and Price (1954). Figure 2 gives the curves for the imaginary part (ϵ'') of the complex relative permittivity, i.e. the dielectric loss factor, for 0·48, 1·0 and 1·5 M solutions in nujol and also a partial curve for a 1·5 M solution in toluene. Davies and Price published a curve for the 0·48 M solution only, but stated that more concentrated solutions also show a resonance absorption.

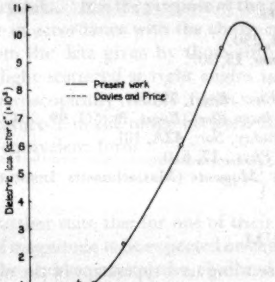


Figure 1.

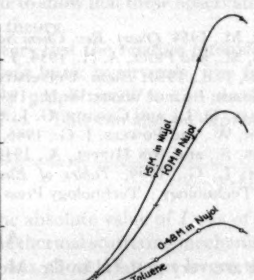


Figure 2.

Referring first to the 0·48 M solution, the results show the presence of a maximum in the loss factor curve at 5 Gc/s. In separate measurements it was found that with decreasing temperature the maximum is displaced to lower frequencies. It is worth noting that with increasing concentration the absorption is displaced slightly towards higher frequencies, corresponding to a marked decrease in the bulk viscosity of the solutions. From these results, and the shape of the curve, it appears that this is a relaxation absorption due to a mechanism of dipole rotation opposed by viscous forces.

The results indicate that there is no resonance absorption maximum in the region of 80 Mc/s and it would appear therefore that Davies and Price have greatly underestimated the experimental errors in their measurements. Further confirmation of these conclusions was obtained from the results for the 1·5 M solution of triethylamine in toluene. This gives very low dielectric loss compared with the same concentration in nujol. If there were a resonance absorption in nujol at these frequencies the solvent would have only a minor effect on the frequency. The absorption observed in the toluene solution is obviously the beginning of the relaxation absorption which is occurring at a higher frequency in the less viscous solvent.

It is, incidentally, possible to obtain a value for the dipole moment μ of triethylamine from these measurements. The absorption is broader than a Debye curve, as is the case for other solutes in nujol (Jackson and Powles 1946, Dryden 1950), but it is possible to calculate μ from the area under the curve according to the following equation, on the assumption that the curve is symmetrical :

$$\mu^2 = \frac{27\boldsymbol{k}T}{2\pi^2 N}\,\frac{\int \epsilon'' d(\log f)}{(\epsilon' + 2)^2}$$

where f is the frequency, ϵ' the relative permittivity, N the number of dipoles/cm^3, $\boldsymbol{k}$ and T are Boltzmann's constant and absolute temperature respectively.

The value obtained, 0·70 Debye unit, can be compared with the mean value of those quoted by Wesson (1948), viz. 0·77 Debye unit.

ACKNOWLEDGMENT

The authors wish to acknowledge the assistance of Mr. P. Buss in the construction of apparatus.

REFERENCES

DAVIES, M., 1954, *Quart. Rev. Chem Soc., Lond.*, **8,** 250.

DAVIES, M., and PRICE, A. H., 1954, *J. Phys. Radium*, **15,** 307.

DRYDEN, J. S., 1950, *Thesis*, University of London.

HARTSHORN, L., and WARD, W. H., 1936, *J. Instn Elect. Engrs*, **79,** 597.

HOLLWAY, D. L., and CASSIDY, G. J. A., 1952, *J. Instn Elect. Engrs*, Pt. III, **99,** 364.

JACKSON, W., and POWLES, J. G., 1946, *Trans. Faraday, Soc.*, **42A,** 101.

ROBERTS, S., and VON HIPPEL, A., 1946, *J. Appl. Phys.*, **17,** 610.

WESSON, L. G., 1949, *Tables of Electric Dipole Moments* (Massachusetts Institute of Technology : Technology Press).

ADDENDUM

We are very grateful to Dr. Meakins for sending us a typescript of the above paper—which came in reply to a letter we addressed him (by surface mail) on 25th July 1955 and from which we quote : "The rather remarkable 'resonance' absorption which we felt we had established for triethylamine, more especially in liquid paraffin, has proved to be the result of marked viscosity changes influencing the solvent absorption and thereby distorting what now appears to be a normal Debye-type absorption." Some weeks before hearing from Dr. Meakins we had submitted a letter to the *Journal of Chemical Physics* indicating that the various criteria we had employed in deducing the unusual 'resonance' form of the absorption were vitiated by a previously unsuspected displacement of the solvent absorption. That note is now in course of publication. It is clear that the above results of Drs. Dryden and Meakins provide an even more complete refutation of our original deduction than we have been able to offer ourselves.

The Edward Davies Chemical Laboratories,
University College of Wales,
Aberystwyth.
12th October 1955

M. DAVIES.
A. H. PRICE.

Temperature Dependence of Coherent Light Scattering in Monocrystals and Intrinsic Block Structure

By R. FÜRTH
Birkbeck College, University of London

MS. received 18*th October* 1955

Some time ago the author pointed out (Fürth 1949) that the concept of an 'intrinsic' block structure of monocrystals, originally put forward by Born, (1947, 1951, 1952) has apparently some bearing on the strength of solids. In a recent publication Humphreys-Owen (1955) has presented new evidence for the existence of a submicroscopic block structure of the predicted order of magnitude from measurements of the wavelength dependence of the intensity of light scattered by NaCl monocrystals; these results have been further confirmed by measurements on quartz crystals which will be published in the near future. Very recently, Taurel and Chapelle (1955) published results on an observed reversible temperature dependence of the total light intensity scattered by NaCl monocrystals. It is the purpose of the present note to show that these observations are also in accordance with the above-mentioned theory.

From the data given by these authors it appears that the relative intensity I of the light scattered at right angles to the incident beam by crystals, free from ultramicroscopically visible faults, is very nearly proportional to the absolute temperature T in the range between 20° and 130°c. This can also be expressed in the equivalent form

$$\alpha = \frac{1}{I}\frac{dI}{dT} = \frac{1}{T}. \qquad \ldots\ldots(1)$$

They further state that for one of their crystals the absolute value of I was of the order of magnitude to be expected on the basis of the thermal scattering mechanism, while for another crystal the observed intensity was about 17 times larger than the value predicted by this theory. In the first case the proportionality of I with T would be in keeping with the theory of thermal scattering, but a reversible temperature dependence of I in the second case can only be explained if it is assumed that the submicroscopic structure, which is responsible for the surplus scattering, also reversibly changes with T.

It follows from the above-mentioned theory of intrinsic block structures that the average number n of atoms along the side a of one of the supposed approximately cubic blocks depends on T according to

$$n = n_0 \tanh(\Theta/2T) \qquad \ldots\ldots(2)$$

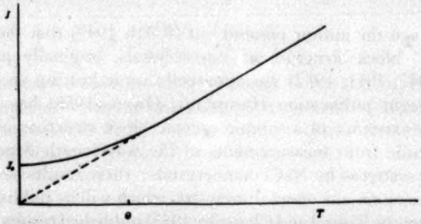
where n_0 is the value of n for $T=0$ and Θ is Debye's characteristic temperature; this shows that for large T/Θ the number n is approximately inversely proportional to the absolute temperature.

Now I will be proportional to the total interface area of the blocks which, in an illuminated region of volume V of the crystal, is equal to $V/a = V/nd$ where d is the interatomic distance. As the latter quantity changes only very slightly with temperature, I should vary inversely proportionally to n for a given V, and thus proportionally to T for sufficiently high temperatures.

For arbitrary T/Θ one has from equation (2)

$$\alpha = -\frac{d(\log n)}{dT} = \frac{1}{T}\,\frac{\Theta/T}{\sinh(\Theta/T)}\,. \qquad \ldots\ldots(3)$$

For NaCl $\Theta = 278°\text{K}$, and thus in the investigated region $\Theta/T \simeq 0{\cdot}8$ and the factor of $1/T$ on the right-hand side of (3) is still practically equal to unity, in agreement with (1). But for very low temperatures α tends exponentially towards zero, and hence we should expect a dependence of I on T of the form shown in the diagram, provided the explanation given is correct.



A more general and rigorous theory of the intrinsic block structure on the basis of the dynamics of crystal lattices will be presented in another publication.

References

Born, M., 1947, *Proc. Math. Phys. Soc. Egypt*, **3**, 35; see also 1951, *Festschr. Akad. Wiss. Gott., Math.-Phys. Kl. 1* and 1952, *Changements de Phases* (Paris: Société de Chimie Physique), p. 334.
Fürth, R., 1949, *Phil. Mag.*, **40**, 1227.
Humphreys-Owen, S. P. F., 1955, *Proc. Phys. Soc.* B, **68**, 325.
Taurel, L., and Chapelle, J., 1955, *C.R. Acad. Sci., Paris*, **240**, 2507.

Some Numerical Evaluations of Heat Flow in the Region Bounded Internally by a Circular Cylinder

By H. GOLDENBERG
Electrical Research Association, Leatherhead, Surrey

Communicated by S. Whitehead; *MS. received 27th July* 1955

§ 1. Introduction

The solutions of many linear heat flow problems have been obtained by operational methods, but in the majority of cases the final solution contains an infinite integral or an infinite series which cannot readily be expressed in terms of analytic or tabulated functions. Consequently the numerical evaluation and publication of such infinite integrals and series becomes of major importance.

The conduction of heat in the region bounded internally by a circular cylinder has received much attention because of its importance in buried pipes and cables (Whitehead 1944), cooling of mines (Smith 1937), and motion in viscous fluids (Batchelor 1954). An excellent description of the principal problems and results obtained until recently has been given by Carslaw and Jaeger (1947, p. 280ff.).

The present note is concerned with this problem subject to the boundary condition that the cylinder surface is maintained at constant temperature while the region outside the cylinder is initially at zero temperature. Theoretical expressions have previously been obtained for the transient temperature at any point outside the cylinder and for the transient flux at the cylinder surface. The latter flux has been evaluated numerically by Jaeger and Clarke (1942) and the result presented graphically by Carslaw and Jaeger (1947). However, no numerical or graphical results appear to have been given for the transient temperatures at individual points outside the cylinder or for the total flux crossing unit area of the surface, up to any time. The object of this paper is the presentation of these results in a manner which should prove useful in a variety of fields.

§ 2. RADIAL HEAT FLOW PROBLEM

Consider the heat conduction problem of the infinite region bounded internally by the circular cylinder $r=a$, the cylinder surface maintained at a constant temperature V, the temperature elsewhere zero initially at time $t=0$.

The temperature $v(r,t)$ at any radius $r>a$ and at any time $t>0$ is given by

$$v(r,t)=V\mathscr{L}^{-1}\left\{\frac{1}{p}\frac{\mathrm{K}_0(r\sqrt{[p/D]})}{\mathrm{K}_0(a\sqrt{[p/D]})}\right\} \qquad \ldots\ldots(1)$$

where K_0 is the modified Bessel function, D is the thermal diffusivity of the region external to the cylinder and

$$\mathscr{L}^{-1}\{F(p)\}=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}[\mathrm{e}^{pt}F(p)]\,dp=\text{inverse Laplace transform of } F(p).$$

$v(r,t)$ has been obtained by Nicholson (1921) using Weber's integral, and by Goldstein (1932) and Carslaw and Jaeger (1940) using operational and Laplace transformation methods. $v(r,t)$ is given by

$$v(r,t)=V+\frac{2V}{\pi}\int_0^\infty \exp(-Du^2t)\frac{[\mathrm{J}_0(ur)\,\mathrm{Y}_0(ua)-\mathrm{Y}_0(ur)\,\mathrm{J}_0(ua)]}{\mathrm{J}_0{}^2(au)+\mathrm{Y}_0{}^2(au)}\frac{du}{u}. \qquad \ldots.(2)$$

This integral does not appear to be expressible directly in terms of analytic or tabulated functions. However, the important case of the transient flux at the surface has been presented graphically by Carslaw and Jaeger (1947, p. 283), and is expressed by

$$\begin{aligned}-k\left(\frac{\partial V}{\partial r}\right)_{r=a}&=\frac{Vk}{\sqrt{D}}\mathscr{L}^{-1}\left\{\frac{1}{\sqrt{p}}\frac{\mathrm{K}_1(a\sqrt{[p/D]})}{\mathrm{K}_0(a\sqrt{[p/D]})}\right\}\\&=\frac{4Vk}{a\pi^2}\int_0^\infty\exp(-Du^2t)\frac{du}{u[\mathrm{J}_0{}^2(au)+\mathrm{Y}_0{}^2(au)]}\\&=\frac{4Vk}{a\pi^2}\,\mathrm{I}(0,1\,;\,T)\end{aligned} \qquad \ldots\ldots(3)$$

where $T=Dt/a^2$ and $\mathrm{I}(0,1\,;\,T)$ is defined as

$$\int_0^\infty \exp(-u^2T)\frac{du}{u[\mathrm{J}_0{}^2(u)+\mathrm{Y}_0{}^2(u)]}.$$

The function $\mathrm{I}(0,1\,;\,T)$ has been tabulated by Jaeger and Clarke (1942) for $T=0{\cdot}01\,(0{\cdot}01)\,1\,(0{\cdot}1)\,10\,(1)\,100\,(10)\,1000$.

From equation (2)

$$\frac{\pi}{2}\left[1-\frac{v(r,t)}{V}\right]=\int_0^\infty\exp(-y^2T)\frac{[\mathrm{J}_0(y)\,\mathrm{Y}_0(\lambda y)-\mathrm{Y}_0(y)\,\mathrm{J}_0(\lambda y)]}{\mathrm{J}_0{}^2(y)+\mathrm{Y}_0{}^2(y)}\frac{dy}{y} \qquad \ldots\ldots(4)$$

where $\lambda = r/a$. In order to evaluate numerically the infinite integral on the right-hand side of this equation for any given value of T the infinite range of integration of y from 0 to ∞ may be limited to a finite range of integration from 0 to y_1 owing to the rapidly decaying nature of the term $\exp(-y^2T)$ with increasing y. In the present instance it is adequate to choose $y_1^2 T \geqslant 10$.

Further as $y \to 0$, $\mathrm{J}_0(y)\,\mathrm{Y}_0(\lambda y) - \mathrm{Y}_0(y)\,\mathrm{J}_0(\lambda y) \to (2/\pi)\ln\lambda$, while the integrand as a whole tends to infinity. For a suitable value of y_1 therefore we may take

$$\begin{aligned}\frac{\pi}{2}\left[1 - \frac{v(r,t)}{V}\right] &= \int_0^{y_1} \exp(-y^2T)\,\frac{[\mathrm{J}_0(y)\,\mathrm{Y}_0(\lambda y) - \mathrm{Y}_0(y)\,\mathrm{J}_0(\lambda y)]}{\mathrm{J}_0^2(y) + \mathrm{Y}_0^2(y)}\,\frac{dy}{y}\\ &= A\int_0^{y_2} \frac{\exp(-y^2T)\,dy}{[\mathrm{J}_0^2(y) + \mathrm{Y}_0^2(y)]y}\\ &\quad + \int_{y_2}^{y_1} \exp(-y^2T)\,\frac{[\mathrm{J}_0(y)\,\mathrm{Y}_0(\lambda y) - \mathrm{Y}_0(y)\,\mathrm{J}_0(\lambda y)]}{\mathrm{J}_0^2(y) + \mathrm{Y}_0^2(y)}\,\frac{dy}{y}\end{aligned}$$

where $\mathrm{J}_0(y)\,\mathrm{Y}_0(\lambda y) - \mathrm{Y}_0(y)\,\mathrm{J}_0(\lambda y)$ is sufficiently close to $(2/\pi)\ln\lambda$ over the range $y = 0$ to $y = y_2$, and A may be taken as

$$A = \tfrac{1}{2}\left[\frac{2}{\pi}\ln\lambda + \mathrm{J}_0(y_2)\,\mathrm{Y}_0(\lambda y_2) - \mathrm{Y}_0(y_2)\,\mathrm{J}_0(\lambda y_2)\right].$$

We therefore have

$$\begin{aligned}\frac{\pi}{2}\left[1 - \frac{v(r,t)}{V}\right] &= A\int_0^{\infty} \frac{\exp(-y^2T)}{\mathrm{J}_0^2(y) + \mathrm{Y}_0^2(y)}\,\frac{dy}{y} - A\int_{y_2}^{y_1} \frac{\exp(-y^2T)}{\mathrm{J}_0^2(y) + \mathrm{Y}_0^2(y)}\,\frac{dy}{y}\\ &\quad - A\int_{y_1}^{\infty} \frac{\exp(-y^2T)}{\mathrm{J}_0^2(y) + \mathrm{Y}_0^2(y)}\,\frac{dy}{y}\\ &\quad + \int_{y_2}^{y_1} \exp(-y^2T)\,\frac{[\mathrm{J}_0(y)\,\mathrm{Y}_0(\lambda y) - \mathrm{Y}_0(y)\mathrm{J}_0(\lambda y)]}{\mathrm{J}_0^2(y) + \mathrm{Y}_0^2(y)}\,\frac{dy}{y}\\ &= A\,\mathrm{I}(0,1;\,T) + \int_{y_2}^{y_1} \frac{\exp(-y^2T)\,[\mathrm{J}_0(y)\,\mathrm{Y}_0(\lambda y) - \mathrm{Y}\mathrm{J}_0(y)\,_0(\lambda y) - A]}{\mathrm{J}_0^2(y) + \mathrm{Y}_0^2(y)}\,\frac{dy}{y}\end{aligned}$$

since for y_1 chosen as above $\displaystyle\int_{y_1}^{\infty} \frac{\exp(-y^2T)}{\mathrm{J}_0^2(y) + \mathrm{Y}_0^2(y)}\,\frac{dy}{y}$ is negligible. We therefore have the formula

$$\frac{v(r,t)}{V} = 1 - \frac{2}{\pi}A\,\mathrm{I}(0,1;\,T) - \frac{2}{\pi}\int_{y_2}^{y_1} \exp(-y^2T)\,\frac{[\mathrm{J}_0(y)\,\mathrm{Y}_0(\lambda y) - \mathrm{Y}_0(y)\,\mathrm{J}_0(\lambda y) - A]}{\mathrm{J}_0^2(y) + \mathrm{Y}_0^2(y)}\,\frac{dy}{y} \quad \ldots\ldots(5)$$

where y_1 is chosen so that range of integration 0 to ∞ of the infinite integral of equation (4) may be reduced to the range 0 to y_1 and y_2 is chosen so that over the range $0 \leqslant y \leqslant y_2$ we may take $\mathrm{J}_0(y)\,\mathrm{Y}_0(\lambda y) - \mathrm{Y}_0(y)\,\mathrm{J}_0(\lambda y) = A$.

The right-hand side of equation (5) has been evaluated numerically for $r = 2a$, $r = 10a$ and $r = 100a$ for a large range of values of the non-dimensional parameter T, and the results are given in figure 1.

The total flux F crossing unit area of the cylinder surface from time $t = 0$ to any time t is obtained by integrating equation (3) with respect to time t and is

$$\begin{aligned}F = \int_0^t -k\left(\frac{\partial V}{\partial r}\right)_{r=a} dt &= \frac{Vk}{\sqrt{D}}\,\mathscr{L}^{-1}\left\{\frac{1}{p^{3/2}}\,\frac{\mathrm{K}_1\,(a\sqrt{[p/D]})}{\mathrm{K}_0\,(a\sqrt{[p/D]})}\right\}\\ &= \frac{4Vk}{a\pi^2}\int_0^t \mathrm{I}\,(0,1;\,T)\,dt = \frac{4Vka}{\pi^2 D}\int_0^T \mathrm{I}(0,1\,;\,T)\,dT.\end{aligned}$$

For small values of T Jaeger and Clarke (1942) give the formula

$$\mathrm{I}\,(0,\,1\,;\,T) = \frac{\pi^2}{4}\left\{\sqrt{\frac{1}{(\pi T)}} + \frac{1}{2} - \frac{1}{4}\sqrt{\frac{T}{\pi}} + \frac{T}{8} - 0{\cdot}1469\,T^{3/2} + 0{\cdot}203\,T^2 - 0{\cdot}315\,T^{5/2} + 0{\cdot}536\,T^3 \ldots\right\}.$$

Hence $$\int_0^T \mathrm{I}\,(0,\,1\,;\,T)\,dT = \frac{\pi^2}{4}\left\{2\sqrt{\frac{T}{\pi}} + \frac{T}{2} - \frac{1}{6\sqrt{\pi}}\,T^{3/2} + \frac{T^2}{16} - 0{\cdot}0588\,T^{5/2} + 0{\cdot}0677\,T^3 - 0{\cdot}09\,T^{7/2} + 0{\cdot}1340\,T^4 \ldots\right\}.$$

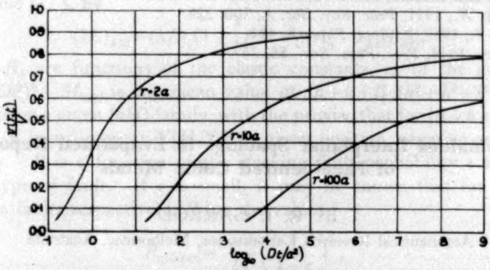


Figure 1. Heat conduction outside the circular cylinder $r=a$, the cylinder surface maintained at constant temperature V, temperature elsewhere zero initially at time $t=0$. Curves giving transient temperatures $v(r,\,t)$ at radii $r=2a$, $r=10a$ and $r=100a$. D is the thermal diffusivity of the region external to the circular cylinder.

While this series expansion is readily applicable for small values of T, it becomes inconvenient for moderate and large values of T, and in fact the tabulated values of $\mathrm{I}\,(0,\,1\,;\,T)$ given by Jaeger and Clarke (1942) were integrated numerically to avoid the use of the series for large values of T. Their upper limit of tabulation $T=1000$ for $\mathrm{I}\,(0,\,1\,;\,T)$ was somewhat extended to permit wider total flux calculation. The total flux F crossing unit area of the cylinder surface from time $t=0$ to any time t has been obtained numerically for a wide range of values of T and is given in non-dimensional form in figure 2.

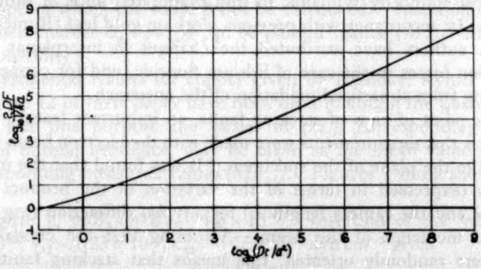


Figure 2. Heat conduction outside the circular cylinder $r=a$, the cylinder surface maintained at constant temperature V, temperature elsewhere zero initially at time $t=0$. Flux F crossing unit area of cylinder surface from time $t=0$ to any time t. D is the thermal diffusivity and k the conductivity of the region external to the circular cylinder.

ACKNOWLEDGMENT

The author wishes to thank the Director of the Electrical Research Association for permission to publish this paper.

REFERENCES

BATCHELOR, G. K., 1954, *Quart. J. Mech.*, **7**, 179.
CARSLAW, H. S., and JAEGER, J. C., 1940, *Proc. Lond. Math. Soc.*, (2), **46**, 361; 1947, *Conduction of Heat in Solids* (Oxford: University Press).
GOLDSTEIN, S., 1932, *Proc. Lond. Math. Soc.*, (2), **34**, 51.
JAEGER, J. C., and CLARKE, M., 1942, *Proc. Roy. Soc. Edinb.* A, **61**, 229.
NICHOLSON, J. W., 1921, *Proc. Roy. Soc.* A, **100**, 226.
SMITH, F. L. P., 1937, *J. Appl. Phys.*, **8**, 441.
WHITEHEAD, S., 1944, *Proc. Phys. Soc.*, **56**, 357.

Anomalous Interplanar Spacings in Evaporated Deposits of Face-centred Cubic Metals

BY R. I. GARROD
Aeronautical Research Laboratories, Melbourne, Australia

Communicated by L. H. Martin; MS. received 15*th August* 1955, *and in amended form* 21*st October* 1955

SEVERAL years ago, Williams (1952), on the basis of precision measurements of electron diffraction transmission patterns, reported small anomalies in the relative interplanar spacings of thin films of copper, silver and gold, prepared by evaporation *in vacuo*. These changes (as compared with the spacings to be expected from a perfect face-centred cubic structure) were ascribed to anisotropic lattice strains resulting from dislocations near to the free surfaces of the globular particles comprising the metal films. Broom (1953, p. 342) has pointed out, however, that Williams' results could alternatively be explained qualitatively in terms of stacking faults (Paterson 1952).

More recently, Halliday, Rymer and Wright (1954) have observed similar effects to those found by Williams, in thin evaporated films of lithium fluoride and copper. In accordance with previous work on gold leaf (Rymer and Butler 1947), these authors have attributed the changes in interplanar spacings to surface-tension forces in the case of lithium fluoride, and for copper to a stress system arising from superficial oxidation of the specimen.

From the point of view of stacking faults, an important feature of the latter experiments is that measurements were made with the electron beam both normal and inclined to the plane of the specimen. It was found that the magnitude of the anomaly (expressed in terms of the variation of the product of electron wavelength λ and the camera length L) for any *hkl* diffraction ring varied with the angle of incidence of the beam. Assuming that the crystallites in the specimen were randomly oriented, this means that stacking faults could not have been the sole cause of the discrepancies observed in copper. On the other hand they could be a major contributory factor. It is therefore of interest to investigate quantitatively the effects produced upon the electron diffraction pattern by a combination of a stress system and stacking faults in the specimen, and to examine the results of Williams and Halliday, Rymer and Wright in the light of such a dual process.

Consider then a face-centred cubic polycrystalline specimen which contains stacking faults and in which there exists a tensile stress p, normal to the specimen plane and at an angle ϕ to the incident electron beam. Assuming that to a close approximation the changes in diffraction ring radius due to stress (Halliday, Rymer and Wright 1954) and to faulting (Paterson 1952) are linearly additive, it follows that the observed value for λL for each *hkl* diffraction ring is related to the true value $(\lambda L)_0$ by

$$(\lambda L)_{hkl} = (\lambda L)_0[1 + \tfrac{1}{2}PM - p(A_\phi + \alpha_{hkl}B_\phi)], \qquad \ldots\ldots(1)$$

where A_ϕ, B_ϕ are functions of the elastic constants, c_{ij} of the material, and $\alpha = \Sigma h^2k^2/(\Sigma h^2)^2$; M_{hkl} is the mean value of $|h+k+l|/(h^2+k^2+l^2)$ for all *hkl* reflections from a given $\{hkl\}$ family, with the proviso that for $|h+k+l| = 3N \pm 1$, $3N$ (where N is an integer), $\Sigma|h|/\Sigma h^2$ is given positive, negative or zero values respectively. P is a function of the stacking fault probability ϵ and depends upon the type of fault. If ϵ is small, it may be shown that for growth and deformation faults respectively, P has the values

$$P_{\text{growth}} = \frac{\sqrt{3}}{2\pi}\left(\frac{\epsilon}{1-\epsilon}\right)^2, \qquad \ldots\ldots(2a)$$

$$P_{\text{deformation}} = \frac{\sqrt{3}}{2\pi}\,\epsilon. \qquad \ldots\ldots(2b)$$

Let $\Delta_\phi \equiv [(\lambda L)_{hkl} - (\lambda L)_{200}]/(\lambda L)_0$ for an angle ϕ, and let Δ_0 be the corresponding quantity for normal incidence of the electron beam.

The following expressions may then be obtained from equation (1):

$$\Delta_\phi - \Delta_0 = \tfrac{3}{2}\,p\alpha_{hkl}B_0\sin^2\phi, \qquad \ldots\ldots(3)$$

$$\Delta_\phi + p\alpha_{hkl}B_\phi = \tfrac{1}{2}P(M_{hkl} - M_{200}) \qquad \ldots\ldots(4)$$

where

$$B_\phi = \left(1 - \frac{3}{2}\sin^2\phi\right)\left(\frac{1}{c_{11}-c_{12}} - \frac{1}{2c_{44}}\right).$$

Hence if $\Delta_\phi - \Delta_0$ is plotted against α_{hkl} a straight line should be obtained from the slope of which the stress p can be calculated. Insertion of this value of p in (4) then enables $\Delta_\phi + p\alpha_{hkl}B_\phi$ to be plotted against $(M_{hkl} - M_{200})/2$ giving the fault probability ϵ.

The experimental results for copper given by Halliday, Rymer and Wright (1954) in figure 12 of their paper have been used to obtain the quantities Δ_ϕ and Δ_0 above. For this purpose the values of $(\lambda L)_0$ corresponding to inclined $(\phi = 40°)$ and normal incidence, may be taken as the mean value in each case of the corresponding measured values $(\lambda L)_{hkl}$. Figures 1 and 2 show the results obtained.

From figure 1, $\tfrac{3}{2}pB_0\sin^2\phi = -4{\cdot}5 \times 10^{-3}$. Insertion of the values $\phi = 40°$, $c_{11} = 1{\cdot}7 \times 10^{12}$, $c_{12} = 1{\cdot}23 \times 10^{12}$, $c_{44} = 0{\cdot}75 \times 10^{12}$ dyn cm^{-2} gives

$$p = -5{\cdot}0 \times 10^9 \text{ dyn cm}^{-2}.$$

From figure 2, $P = 2{\cdot}9 \times 10^{-3}$, corresponding to growth or deformation fault probabilities ϵ of 0·09 or 0·01 respectively.

The inclusion of stacking faults effects makes a significant improvement to the agreement between calculated (d_c) and observed (d_0) values of interplanar spacings. For example, the above results lead to a value for $\langle[(d_c - d_0)/d_0]^2{}_{av}\rangle^{1/2}$ of $0{\cdot}8 \times 10^{-4}$, compared with a value of $1{\cdot}9 \times 10^{-4}$ if elastic stresses alone are considered.

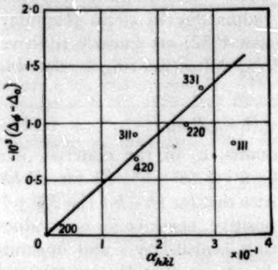


Figure 1. Use of the experimental results of Halliday, Rymer and Wright (1954) for copper to obtain the value of the stress p in the specimen.

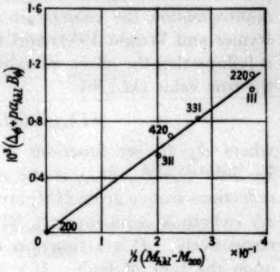


Figure 2. Results of Halliday, Rymer and Wright (1954) plotted to obtain the stacking fault probability ϵ in their copper specimen.

Growth faults can arise as the evaporated film is formed, but deformation faults require a plastic strain in the specimen during or subsequent to its formation. Murbach and Wilman (1953) have shown that stresses are present in evaporated metals films of the order of 1000–3000 Å in thickness and that in several cases, including copper, the stress exceeds the elastic limit of the material. Furthermore, Greenough and Smith (1955) have inferred from x-ray evidence that deformation faults are present in cold-worked copper filings ($\epsilon = 7 \times 10^{-3}$).

The value obtained here, however, shows that if deformation faults exist in the evaporated copper film, they occur on the average approximately once every 200 Å in the [111] direction. Since the crystallite size is unlikely to exceed this order of magnitude, the faults are predominantly of the growth type.

Finally, the line broadening due to stacking faults needs to be considered. From measurements of diffraction ring breadths and the fact that the 200 ring was appreciably broader than the other reflections, Halliday, Rymer and Wright concluded that the crystallites in their copper films were in the form of discs parallel to (111) planes with a diameter and thickness of order 90 Å and 15 Å respectively. The above results on the other hand indicate that growth faults occur on the average once every eleven (111) layers (23 Å) in each crystal. Consideration of the radial broadening due to faulting shows that for the first four reflections the integral breadths due to this effect alone should increase in the order 311, 111, 220, 200, which is roughly in agreement with the experimental results of Halliday, Rymer and Wright, although their interpretation was different. It seems probable therefore that growth faults are making a major or even prime contribution to the observed radial broadening. In consequence it is suggested that the figure of 15 Å given previously for the average crystallite thickness is well below the true value.

With reference to the results of Williams (1952), equations (3) and (4) above cannot be applied owing to the availability of data only for normal incidence of the beam. For this case however, by analogy with (4),

$$\Delta_0 = P(M_{hkl} - M_{200})/2 - p\alpha_{hkl}B_0. \qquad \ldots\ldots(5)$$

Thus from the six values for $(\lambda L)_{hkl}$ given in figure 1 of the paper by Williams, five numerical equations are available containing the two unknown parameters p and P. Applying a least squares method (Whittaker and Robinson 1937, p. 210) to the data gives a value for $p = -3{\cdot}7 \times 10^9$ dyn cm^{-2} and for the growth fault probability $\epsilon = 0{\cdot}06$. These values are thus comparable in magnitude with the corresponding quantities for copper deduced above from the results of Halliday, Rymer and Wright.

ACKNOWLEDGMENT

This Note is published with the permission of the Chief Scientist, Department of Supply, Australia.

REFERENCES

BROOM, T., 1953, *Properties of Metallic Surfaces* (London : Institute of Metals, Monograph & Report Series No. 13).

GREENOUGH, G. B., and SMITH, E. M., 1955, *Proc. Phys. Soc.* B, **68,** 51.

HALLIDAY, J. S., RYMER, T. B., and WRIGHT, K. H. R., 1954, *Proc. Roy. Soc.* A, **225,** 548.

MURBACH, H. P., and WILMAN, H., 1953, *Proc. Phys. Soc.* B, **66,** 905.

PATERSON, M. S., 1952, *J. Appl. Phys.*, **23,** 805.

RYMER, T. B., and BUTLER, C. C., 1947, *Proc. Phys. Soc.*, **59,** 541.

WHITTAKER, E. T., and ROBINSON, G., 1937, *The Calculus of Observations* (London Blackie).

WILLIAMS, E. C., 1952, *Research, Lond.*, **5,** 392.

REVIEWS OF BOOKS

Frontiers of Astronomy, by FRED HOYLE. Pp. xvi + 360. (London: Heinemann, 1955). 25*s*.

Theory in astronomy has a dual role. Not only must it present a quantitative description of observation in terms of terrestrial physics, but on the frontiers where we are almost in complete ignorance it must simulate experiment by predicting the observational consequences of models in which one or more of the possible mechanisms are dominant. A classical example of the latter role is Schwarzschild's prediction of the darkening of the limb of the sun for a photosphere in convective or radiative equilibrium. In both these aspects of his work the theorist must further recognize that the observational concepts no longer have the simple content which they have in terrestrial physics. Thus when we say the Andromeda Nebula is $1{\cdot}4 \times 10^{24}$ cm away, the distance so quoted contains within it a wealth of empirical observational laws and a chain of hypotheses. We may hope that this distance is the same as would be obtained from repeated applications of a rigid measuring rod, but the recent doubling of the distance-scale for extra-galactic objects shows how naïve such a hope must be. The problems involved in the observational concepts of astronomy are somewhat subtle ones, and the theorist has to step warily both in their use and in the prediction of the observational consequences of theoretical models.

Above all the astronomical theorist needs physical insight of the highest order. In this respect few are more gifted than Mr. Fred Hoyle, but *Frontiers of Astronomy*, in which he presents his latest views on stars and the universe, to say nothing of the earth and the solar system, will be read with disappointment by astronomers. Nor can it be recommended to the physicist as an accurate description of the present state of astronomical knowledge. Not only does it contain almost no account of the transfer of radiation in stellar photospheres, so vital in measuring the rate of energy generation as well as being one of the few branches of astrophysics where precise quantitative tests of theory are possible, but also the account in Chapter 10 of distance measurement, quite basic in all Hoyle's theories, can only be described as a travesty of the facts. How such a chapter could be written without even a mention of Hertzsprung, the founder of the photometric and spectroscopic methods of distance measurement, or of Trumpler and the first quantitative determination of interstellar absorption, almost defies explanation!

What the book does contain are Hoyle's views on the observationally inaccessible interiors of stars, the life histories of the stars if these views are correct and the consequent description of the universe. These topics can be described as on the frontiers of astronomy, indeed on the very far frontiers, but whether these are the frontiers from which any substantial advance is likely in our present ignorance of most of the facts is a question on which opinions will naturally differ. There is indeed something to be said for a theoretical attack in advance of quantitative observation since the theory suggests the observations which can support or disprove it. But there seems nothing whatever to be said for attempting to present the highly intuitive and necessarily speculative theories involved in such an attack in a book clearly designed for a reader with little physics and less astronomy. The validity of such theories can only be assessed by professional workers in the field, and by them only if the basic hypotheses are clearly set forth together with some of the mathematical apparatus of their development. What is essential, *pace* the author, is the kill, not the excitement of the chase, nor is anything gained by following his technique of moving "helter-skelter to a dramatic conclusion". H. H. PLASKETT.

X-Ray Diffraction by Polycrystalline Materials, edited by H. S. PEISER, H. P. ROOKSBY and A. J. C. WILSON. Pp. 725. (London: Institute of Physics, Physics in Industry Series, 1955.) 63*s*.

This book has apparently been written with two objects in view; to provide a standard reference book on the Powder method, and subsidiarily to provide propaganda for the use of such methods in industry. It has a considerable weight of authority having been sponsored by the X-Ray Analysis Group of the Institute of Physics and carrying a commendatory foreword by Sir Lawrence Bragg.

The co-operation of 31 authors has enabled a comprehensive and thorough treatment to be made, and the editors have largely avoided the vices of repetition and overlap of subject matter. On the other hand, the arrangement is in places arbitrary and there are considerable differences of style and in level of approach.

The first part deals with experimental technique, that is largely with cameras, although there is a chapter on X-Ray Generators and Film Measuring Equipment and another on Counter Diffractometers. American practice is considered but nevertheless the emphasis is on methods and equipment of European origin. Besides the usual Debye–Scherrer arrangement with a fine rod specimen, focusing cameras and the various kinds of special purpose cameras are fully described. It is difficult to find fault with these descriptions, but particularly where the apparatus concerned is not readily available either commercially or otherwise, the authors would have helped many readers by giving also a fuller assessment of the value of some of the methods proposed.

The second part covers interpretation, including identification, indexing lattice parameters, intensities, line profiles, and preferred orientation. The discussion is full and easy to follow and working procedures are given for many kinds of calculation. The first chapter of the book which is a summary of lattice theory and x-ray optics belongs more naturally to this part than to the first. In contrast to the remainder of the book it is rather abstract in style, and it is a pity as regards the industrial reader to commence in this way.

The third part consists of nine essays on the application of the powder method in various fields of research, with a final chapter on photography which has strayed from part one. Some of these essays develop further points made in the first two parts but mostly they are progress reports in various specialized fields. There is considerable variation in the kind of treatment given but it is necessarily not nearly so full as in the other parts, nor is there much sense of unity in the subject matter. Only those aspects involving the x-ray results are shown and it is difficult to appreciate fully the situation in these fields without the general view.

There is finally a useful appendix giving several tables (some new in content or arrangement) an extensive bibliography related to chapters, and author and subject indices.

To summarize, the book provides an excellent laboratory manual on the X-ray powder method. A good deal of the ground has been covered before, but the treatment now given is altogether more complete. It will be of constant value to anyone using powder or fibre patterns, and occasionally no doubt to other X-ray workers as well. D. E. BROMLEY.

Theory of Lenses, by E. W. H. SELWYN. Pp. 62. (London: Institute of Physics, 1955.) 5*s*.

In this book, which is one of the series of monographs published by the Institute of Physics for general reading by students in courses at the level of the Higher National Certificate in Applied Physics, the author has been successful in compressing much information into a relatively small space. The book is not above criticism: there are a few faults the correction of which would, in the opinion of the reviewer, enhance its value.

There are four chapters. The first, entitled 'Theory of Light' (more appropriately 'Wave Theory of Light'), deals with the propagation of light on the basis of wave theory. Since space is apparently at a premium the consideration of a doublet source to provide an explanation of the non-existence of a back-wave is hardly necessary: it would be sufficient to quote the obliquity factor, and to mention that Kirchhoff first gave the exact formulation of Huygens' principle. (It is welcome to see Huygens' name spelt correctly !)

The first section of Chapter 2 logically belongs to Chapter 1 as it deals with physical optics. This section describes the formation of the image produced by a spherical concave wave front—avoiding the Bessel functions in Airy's analysis by the use of vectors. This is very good. Discussion of resolving power should follow the formula for the radius of the first dark ring.

The second part of Chapter 2 contains the 'First Approximation' which is the title of the chapter. The section headed 'Ray Theory' is confusing. N_1 is used (line 3, page 29) without being defined, and (line 4, same page) U_1 appears to be a misprint for N_1. At the end of the same section a formula is given which includes f and f', followed by the comment "All this would be very satisfactory if we knew what f and f' were. To find out we must go back to physics." Unfortunately f and f' do not reappear in the book so the reader is left to presume that all this is not satisfactory (or else that a section has been omitted from the chapter).

In the third chapter Seidel aberrations, the second approximation, are very clearly discussed, and there is a useful section on Brightness of Images.

The last chapter discusses correction for colour and high order aberrations. The specification of glass by its index and dispersion or v-number should be mentioned here (the v-number is specified but not named). The discussion of equation (4.5) is too condensed for clarity: it appears to imply that differing refractive indices in the two components of an achromatic lens is a condition for stability of chromatic correction.

The final section entitled 'Real Lenses' is a capable piece of concise writing. It describes, with an admirable little diagram, most of the forms of optical systems in general use.

This could well become a useful little book.

G. W. HAMSTEAD.

University Physics, by F. Sears and M. W. Zemansky. 2nd Edn. Pp. viii + 1031. (Cambridge, Mass.: Addison-Wesley, 1955.) $10.

This book, the authors say, " is based on Sears' three volume work *Principles of Physics*. . . . Those topics which fall within the scope of intermediate physics have been eliminated. . . . It is intended for students of science and engineering who are taking a course in calculus concurrently and to whom calculus is still a new tool ".

The book is well written, but it aims at a far lower level than is usual in English honours courses. Some detailed criticism will make this clear.

The discussion of the first law of thermodynamics is excellent, but only thirteen pages are devoted to the second law, the identity of the Kelvin and Clausius formulations of this law is stated without proof, and the concept of entropy is not introduced.

The development of the magnetism section is based on current elements. This is something of a departure in so elementary a text, and is very successful ; it is a pity that free magnetic poles are introduced later on, but this is probably inevitable at this level of instruction. In view of the sophisticated logical structure, it was rather surprising to find a ' proof ' of Faraday's law of induction on p. 602, which the reviewer would like to have seen more critically discussed.

The pattern produced by a diffraction grating is not derived. Only the position of the principal maxima is discussed, and no adequate reference is made to the resolving power of this instrument. Unpolarized light is rather ambiguously described on p. 865 as " a mixture of waves linearly polarized in all possible transverse directions."

The reviewer can find no mention of the method of electrostatic images, the thermodynamics of an electric cell, group velocity, thick lens theory, Brewster's fringes, and the Michelson and Fabry–Perot interferometers. The mathematical level is, as the authors admit, low ; thus formal vector notation is avoided, and the discussion of rainbows merely says (p. 740) " Descartes computed the paths of some thousands of rays incident at different points on the surface of a raindrop ", without deriving the maximal condition by calculus.

A student who relied solely on this book would be insufficiently prepared for, say, Part 1 of the Natural Sciences Tripos. Nevertheless, within this rather severe limitation, the book is attractively written, and contains stimulating examples. It might be profitably read in conjunction with other textbooks. The price is very high.

A. V. COHEN.

CONTENTS OF SECTION A

PAGE

Mr. I. F. Wright, Mr. D. R. O. Morrison, Mr. J. M. Reid and Mr. J. R. Atkinson. The Photodisintegration of Nitrogen 77

Prof. E. A. Stewardson and Mr. J. E. Wilson. The $M_{IV,V}$ Spectra of ^{68}Er . 93

Mr. J. K. Crawshaw and Dr. H. Elliot. Search for a Sidereal Variation of the Cosmic Ray Intensity at High Energies 102

Dr. E. P. George. Observations of the Nuclear Interactions of a 315 mev Bremsstrahlung γ-ray Beam 110

Mr. I. S. Hughes and Mr. D. Sinclair. The Polarization of γ-Radiation from the Reactions ^{27}Al(p, γ)^{28}Si, ^{26}Mg(p, γ)^{27}Al, ^{23}Na(p, γ)^{24}Mg 125

Dr. E. H. S. Burhop. The Effect of Finite Nuclear Size on Electron Pair Production 135

Mr. K. R. Allen, Mr. M. M. Lipsicas, Mr. D. Major and Mr. K. Phillips. The Scattering of Fast Charged Particles—I : On the Single Scattering of 13·5 mev Electrons in Xenon 141

Prof. P. B. Moon. Space Charge and Ionization Phenomena in Constant-gradient Proton Synchrotrons 153

Mr. B. R. Judd. The Structures of the Ground Multiplets of Certain Rare-Earth Ions 157

Dr. D. St. P. Bunbury, Prof. S. Devons, Dr. G. Manning and Mr. J. H. Towle. Measurement of γ-Transition Lifetimes by Recoil Methods—II : M1 Transitions in ^{7}Li and ^{7}Be 165

Prof. S. Devons, Dr. G. Manning and Mr. J. H. Towle. Measurement of γ-Transition Lifetimes by Recoil Methods—III : E2 Transitions in some Even–Even Nuclei 173

Research Notes :

- Dr. R. F. Barrow and Dr. A. R. Downie. The Identification of a Band System, B($^2\Sigma^+$) $-$A$^2\Sigma^+$, in OH and OD 178
- Dr. K. Venkateswarlu and Mr. S. Sundaram. Evaluation of Force Constants from Raman Effect Data : Molecules, Radicals and Groups of Pyramidal XY_3 Type 180

Letters to the Editor :

- Dr. S. L. Altmann. Spherical Harmonics with the Symmetry of the Close-Packed Hexagonal Lattice 184
- Dr. B. Schiff. The Cellular Method for Close-Packed Hexagonal Titanium 185

Reviews of Books 187

Contents of Section B 188

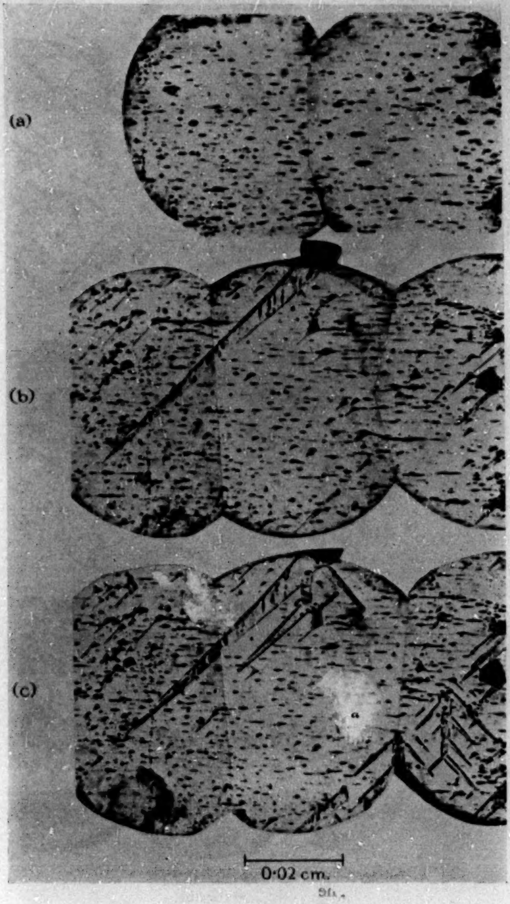


PLATE I.

Figure 1. Series recording nucleation and demagnetization in (001) surface as applied field in [100] direction, i.e. from bottom to top of pictures, is decreased from (*a*) to (*f*). The effective field for (*a*) is about 200 Oe, for (*b*) 150 Oe, for (*c*) 70 Oe.

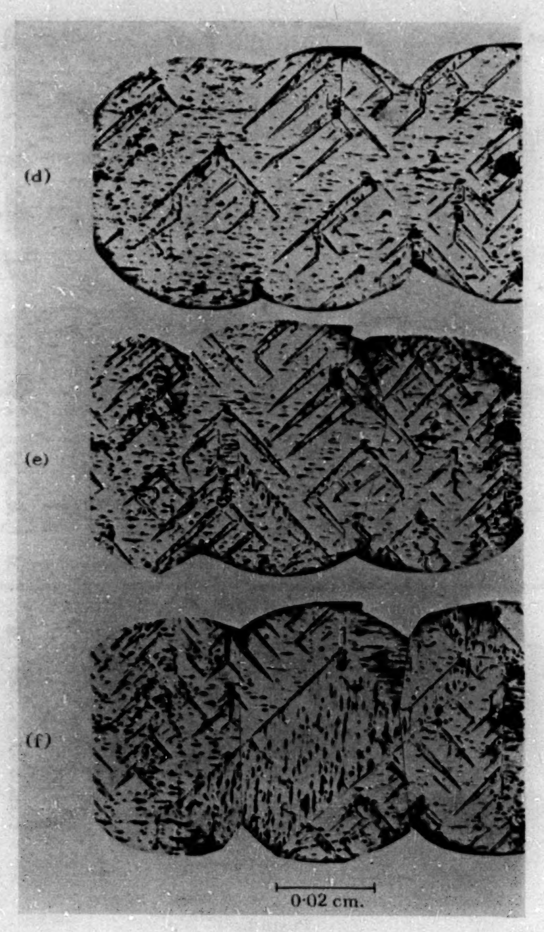


PLATE II.
Figure 1 (*cont.*).

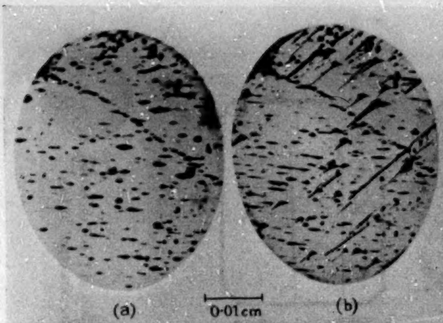


FIG. 2

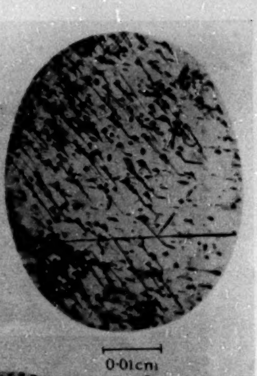


FIG. 3

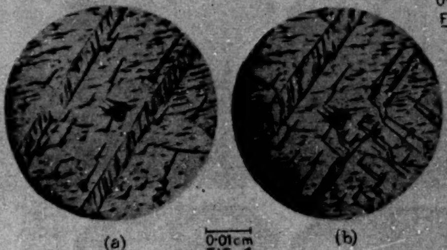


FIG. 4

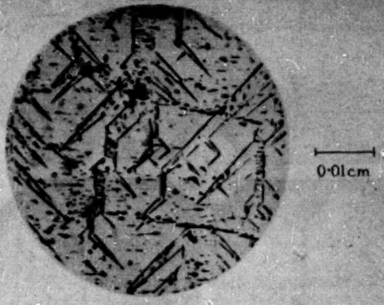


FIG. 5

PLATE III.

Figure 2. (*a*) Surface defects on saturated crystal surface. (*b*) Spike transverse domains on this surface when field is reduced to 150 Oe.

Figure 3. Spike domains on a badly pitted portion of the surface.

Figure 4. (*a*) Spike transverse domains; (*b*) appearance of reverse domains following slight reduction in field from (*a*).

Figure 5. Illustrating close association of reverse and transverse domains.

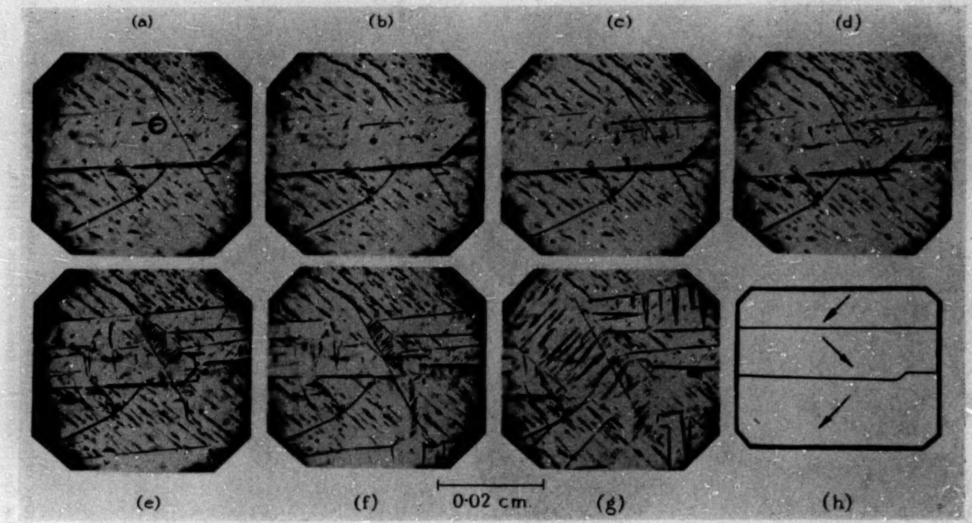


PLATE IV.

Figure 6. Series recording nucleation and demagnetization in (001) surface as applied field in [110] direction, i.e. from top to bottom of pictures, is decreased from (*a*) to (*g*). Diagram (*h*) illustrates structure in (*a*).

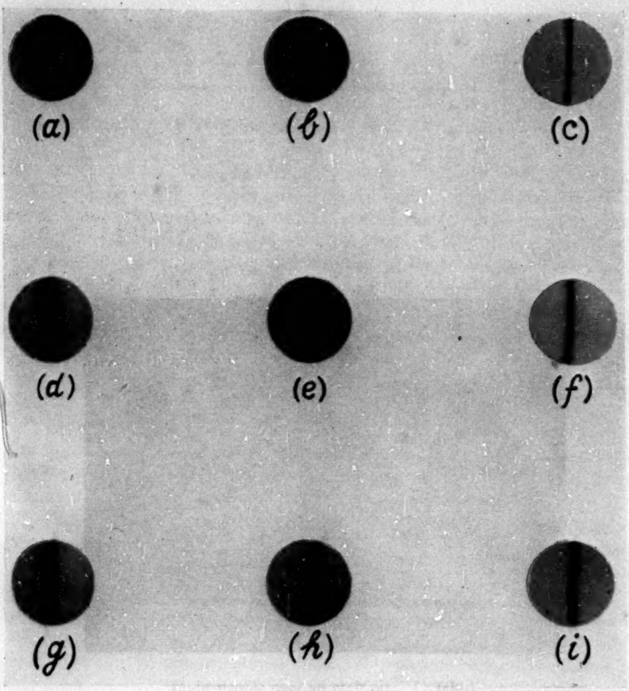
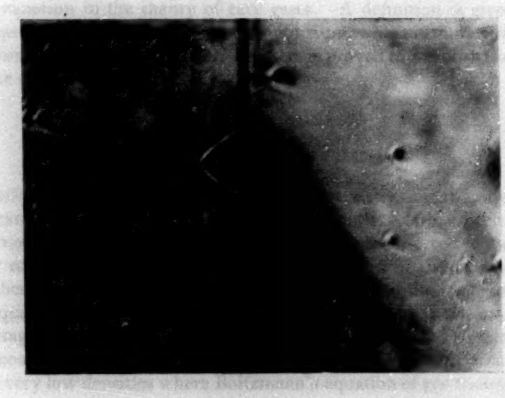


Figure 2.

(*a*) Cu film deposited at -183°c; photographed at -183°c. (200) reflection.
(*b*) Same Cu film as in (*a*) but brought to room temperature. (200) reflection.
(*c*) Cu film as in (*a*) and (*b*) after annealing at 400°c for two hours. (200) reflection.
(*d*) Cu film deposited at -183°c and annealed to room temperature in oxygen. (200) reflection.
(*e*) Cu film deposited at -183°c and annealed to room temperature after the admission of a small amount of oxygen; photographed after three days. (200) reflection.
(*f*) Cu film as in (*e*) stabilized with excess oxygen after 24 hours at room temperature. (200) reflection.
(*g*) Cu film deposited on annealed Cu at -183°c and annealed to room temperature in oxygen. (200) reflection.
(*h*) Cu film deposited at room temperature. (311) doublet.
(*i*) Cu film deposited at room temperature and immediately exposed to air. (311) doublet imperfectly resolved.

PROC. PHYS. SOC. VOL. 69, PT. 2—B (V. G. BHIDE, J. N. DAS AND P. V. KHANDEKAR)



Light profile over an etch pit. (×600.)

Molecular Theory of Irreversible Processes in Fluids

By H. S. GREEN
University of Adelaide, South Australia

MS. received 21st March 1955 and in final form 17th November 1955

Abstract. An attempt is made to formulate a theory of irreversible processes in fluids with the same degree of completeness and rigour as the theory of equilibrium based on statistical mechanics. The theory of the grand partition function is generalized to systems not in thermodynamic equilibrium, and to determine the distribution functions an equation is derived which is strictly analogous to Boltzmann's equation in the theory of rare gases. A definition is given of the entropy of a system not in thermodynamic equilibrium, which is more satisfactory than others previously proposed. An *H*-theorem is proved to establish the natural increase of the entropy so defined.

---

§ 1. Introduction

A certain amount of progress has been made recently towards a qualitative understanding of irreversible processes in fluids. It has been realized that momentum and energy transfer are generally effected in two ways: by the direct motion of the molecules, and by the action of the intermolecular forces. Of these, the former is dominant in rare gases, and the latter in liquids. But on the quantitative side the theory of irreversible processes has not yet reached the stage where a well-defined procedure can be stated, by which coefficients of viscosity and thermal conduction could, in principle, be calculated. Except at the very low densities where Boltzmann's equation of gas theory applies, the theory does not, in fact, approach the completeness of the statistical theory of thermodynamic equilibrium.

The detailed theories of Kirkwood and his collaborators (Kirkwood 1946, Kirkwood, Buff and Green 1949, Irving and Kirkwood 1950, Zwanzig, Kirkwood, Oppenheim and Alder 1954) and Born and Green (1946, 1947) have both met with difficulties of a fundamental nature at the point where irreversibility should enter the theory. The mechanical laws, it is well known, are time-reversible and fail to indicate which of the infinite variety of physically possible states of a molecular assembly are favoured in real irreversible processes. In the kinetic theory of gases, the author (Green 1952, Ch. 8) has shown that an additional statistical postulate serves this purpose. For condensed systems, on the other hand, Kirkwood made a series of approximations suggested by the theory of a Brownian particle. Even then, the difficulty remains that the linear differential equations obtained require for their solution a boundary condition not supplied by the theory. Additional assumptions have to be made which (see Green 1952, pp. 156–160) are practically equivalent to the adoption of a macroscopic molecular model. In the work of Born and Green a difficulty was found in closing the infinite chain of equations which connect the various distribution functions.

Some authors (e.g. Klein and Prigogine 1953) have mistakenly supposed that the solution proposed for this difficulty, the superposition approximation, was intended also to supply the element of irreversibility in the theory. In fact the superposition approximation was used only for thermodynamical equilibrium and, to determine the steady state of a non-uniform fluid, it was assumed that this state depends only on the local values of the macroscopic parameters (density, temperature and fluid velocity) and their space derivatives. Unfortunately this assumption alone did not suffice to determine a unique solution of any problem, and to obtain an explicit formula for the coefficient of viscosity, Born and Green had to 'guess' the deformation of the molecular structure of the fluid by the flow.

The difficulty just mentioned can be related to another difficulty, concerning the definition in molecular terms of the entropy of a fluid which is not in equilibrium. If one adopts the most obvious definition, one finds that the entropy of the fluid can only increase by virtue of its interaction with external systems. Some authors (e.g. Davydov 1947) have accepted and tried to rationalize this result. As classical thermodynamics specifies only the sign, and not the magnitude of the change of entropy in irreversible processes, their standpoint is perfectly tenable. But others (see Tolman 1938, §51) have preferred to modify the definition of entropy by introducing 'coarse-grained' densities in phase space. Each of these alternatives has its disadvantages, and the fact that a rigorous *H*-theorem is lacking for condensed fluids, regarded as isolated systems, has handicapped the development of a detailed theory of irreversible processes, even though the general theory of the Gibbs ensemble is known (see Cox 1950).

In this paper, the author will outline a solution of these problems, helped by the introduction of a new set of velocity distribution functions. The equilibrium form of these functions arises naturally in the theory of the grand partition function, applied to fluids, and they have been studied in some detail by Mayer (1942). In their generalization to non-uniform fluids, they allow one to define the entropy in a precise, realistic way, and to formulate a theory of irreversible processes in dense fluids with the same degree of rigour as Boltzmann's theory of rare gases. In fact, the resulting theory contains as special cases both gas theory and the theory of the grand partition function. In particular, irreversibility is introduced by a generalization of the statistical postulate of gas theory. But the computational difficulties involved in the exact determination of coefficients of viscosity and thermal conduction for liquids are not less than those in evaluating the partition function and solving Boltzmann's equation combined. What can be expected, beyond a solution in principle of these difficulties, is some insight into which approximations are reasonable and likely to give good results. One can also expect to compute, without great difficulty, corrections for dense gases to the results of gas theory.

§2. DISTRIBUTION FUNCTIONS FOR A FINITE DOMAIN

Attention will be restricted to a simple fluid in which the local molecular number density n, temperature T and velocity of flow $\mathbf{u}$ are prescribed throughout the fluid. The notation is a slight condensation of that used previously (Green 1952) by the author: $d\mathbf{x}_q$ represents the product of the volume elements $d\mathbf{x}^{(1)}, \ldots d\mathbf{x}^{(q)}$, and $d\boldsymbol{\xi}_q$ the product of the elementary velocity ranges $d\boldsymbol{\xi}^{(1)}, \ldots d\boldsymbol{\xi}^{(q)}$;

$d\tau^{(q)}$ means $d\mathbf{x}^{(q)}\, d\boldsymbol{\xi}^{(q)}$ and $d\tau_q$ means $d\mathbf{x}_q d\boldsymbol{\xi}_q$. Then $n_q d\mathbf{x}_q$ is the probability that each of the volume elements $d\mathbf{x}^{(1)}, \ldots d\mathbf{x}^{(q)}$ is occupied by (the mass-centre of) some molecule; and $f_q d\tau_q$ represents the probability that these volume elements are occupied by molecules with velocities $\boldsymbol{\xi}^{(1)}, \ldots \boldsymbol{\xi}^{(q)}$ in the ranges $d\boldsymbol{\xi}^{(1)}, \ldots d\boldsymbol{\xi}^{(q)}$. If Φ_q is the total potential energy of a group of q molecules at $\mathbf{x}^{(1)}, \ldots \mathbf{x}^{(q)}$, including their mutual potential energies but not their energies of interaction with other molecules,

$$\frac{d_q f_q}{dt_q} = \frac{1}{m} \int \sum_{i=1}^{q} \frac{\partial \Psi_{q+1}}{\partial \mathbf{x}^{(i)}} \frac{\partial f_{q+1}}{\partial \boldsymbol{\xi}^{(i)}}\, d\tau^{(q+1)} \qquad \ldots\ldots(1)$$

where

$$\frac{d_q}{dt_q} = \frac{\partial}{\partial t} + \sum_{i=1}^{q} \left(\boldsymbol{\xi}^{(i)} \cdot \frac{\partial}{\partial \mathbf{x}^{(i)}} - \frac{1}{m} \frac{\partial \Phi_q}{\partial \mathbf{x}^{(i)}} \cdot \frac{\partial}{\partial \boldsymbol{\xi}^{(i)}} \right) \qquad \ldots\ldots(2)$$

and

$$\Psi_{q+1} = \Phi_{q+1} - \Phi_q. \qquad \ldots\ldots(3)$$

Assuming the molecules interact only in pairs, $\Phi_q = \Sigma_{j>i} \phi^{(ij)}$, where $\phi^{(ij)}$ is the interaction energy of molecules at $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(j)}$. The functions f_q and n_q are related by

$$n_q = \int f_q d\boldsymbol{\xi}_q \qquad \ldots\ldots(4)$$

and in thermodynamic equilibrium only one has

$$f_q = (\beta m / 2\pi)^{3q/2} \exp\left(-\tfrac{1}{2}\beta m \sum_{i=1}^{q} \boldsymbol{\xi}^{(i)2}\right) n_q \qquad \ldots\ldots(5)$$

where $\beta = (\dot{k}T)^{-1}$.

A new set of velocity distribution functions F_q will now be introduced, with special reference to a compact finite domain of the fluid, with volume V. Generally this volume should be large enough to allow one to neglect surface effects, but the only other important restriction is that its surface S should move everywhere with the local velocity $\mathbf{u}$ of the fluid. The function F_q is defined so that $F_q d\tau_q$ is the probability of finding the volume elements $d\mathbf{x}^{(1)}, \ldots d\mathbf{x}^{(q)}$ within V occupied by molecules with velocities in the specified ranges, *and no other molecules within V*. All space integrals will in future be supposed to extend only over the volume V. The functions f_q are obviously related to the F_q by means of

$$\begin{aligned} f_q &= F_q + \int F_{q+1} d\tau^{(q+1)} + \tfrac{1}{2} \iint F_{q+2} \ldots \\ &= \sum_{j=0}^{\infty} (j!)^{-1} \int \overset{(j)}{\ldots} \int F_{q+j} d\tau^{(q+1)} \ldots d\tau^{(q+j)}. \end{aligned} \qquad \ldots\ldots(6)$$

Replacing q by $q+i$ in this relation, and integrating over the phase space elements $d\tau^{(q+1)}, \ldots d\tau^{(q+i)}$, one obtains for $i = 0, 1, 2$ etc. a set of equations from which the F_q can be determined:

$$F_q = \sum_{j=0}^{\infty} (-1)^j (j!)^{-1} \int \overset{(j)}{\ldots} \int f_{q+j} d\tau^{(1)} \ldots d\tau^{(j)}. \qquad \ldots\ldots(7)$$

It should be noticed that the infinite series appearing in (6), (7) and subsequent equations are always convergent, since it is very improbable that a number of molecules far in excess of the mean number $\langle q \rangle$ will be found in V. This has been pointed out by Mayer (1942) and Kirkwood and Salsburg (1953). The maximum degree of convergence is attained when V is as small as possible, consistent with the other requirements of the theory. In the case of a rare gas, one can always choose V so that the first term of the series is a good approximation to the sum. For a condensed system, however, it is to be expected that the main contribution to the sum will come from that part of the series in the neighbourhood of the $\langle q \rangle$th term.

A set of distribution functions N_q, bearing the same relation to n_q as F_q to f_q, can be defined, with the property

$$N_q = \int F_q d\boldsymbol{\xi}_q \qquad \ldots\ldots(8)$$

and such that

$$P_q = (q!)^{-1} \int N_q d\mathbf{x}_q \qquad \ldots\ldots(9)$$

represents the probability of finding precisely q molecules in V. If P_0 is the probability of finding V empty, one must have

$$\sum_{q=0}^{\infty} P_q = 1. \qquad \ldots\ldots(10)$$

The average number of molecules to be found in V is

$$\langle q \rangle = \sum_{q=0}^{\infty} qP_q = \int f_1 d\tau_1 \qquad \ldots\ldots(11)$$

and the mean square number

$$\langle q^2 \rangle = \sum_{q=0}^{\infty} q^2 P_q = \int f_2 d\tau_2 + \int f_1 d\tau_1. \qquad \ldots\ldots(12)$$

In thermodynamic equilibrium, the values of the N_q and F_q can be written down with the help of statistical mechanics (cf. Mayer 1942). Since the work required to evacuate the volume V is pV, where p is the hydrostatic pressure,

$$F_0 = N_0 = P_0 = \exp(-\beta pV). \qquad \ldots\ldots(13)$$

Also, if z is the activity, $N_1 = zN_0$; and in general

$$N_q = z^q N_0 \exp(-\beta \Phi_q). \qquad \ldots\ldots(14)$$

Hence

$$F_q = F_0 \exp\{\beta(q\chi - E_q)\} \qquad \ldots\ldots(15)$$

where

$$\chi = \beta^{-1}\{\ln z + \tfrac{3}{2}\ln(\beta m/2\pi)\} \qquad \ldots\ldots(16)$$

is Gibbs' thermodynamic potential per molecule and

$$E_q = \Phi_q + \tfrac{1}{2}m \sum_{i=1}^{q} \boldsymbol{\xi}^{(i)2}. \qquad \ldots\ldots(17)$$

These last results are based implicitly on the neglect of surface interactions, and so do not hold exactly if one of the positions $\mathbf{x}^{(1)} \ldots \mathbf{x}^{(q)}$ is on or near the surface S. Later an expression will be required for F_{q+1}, where the point $\mathbf{x}^{(1)}$, say, is on S, and the velocity $\boldsymbol{\xi}^{(1)}$ is directed inwards so that a molecule there can have had little or no interaction with others at $\mathbf{x}^{(2)}, \ldots \mathbf{x}^{(q+1)}$. Let $F_q^{[1]}$ denote the function F_q of the variables $\mathbf{x}^{(2)}, \ldots \mathbf{x}^{(q+1)}$; $\boldsymbol{\xi}^{(2)}, \ldots \boldsymbol{\xi}^{(q+1)}$. Assuming $F_{q+1} = \phi F_q^{[1]}$, where ϕ depends only on $\mathbf{x}^{(1)}$ and $\boldsymbol{\xi}^{(1)}$, it follows from (6) that

$$f_1 = \sum_{j=0}^{\infty} (j!)^{-1} \int^{(j)} \cdots \int \phi F_j^{[1]} d\tau^{(2)} \ldots d\tau^{(j+1)} = \phi \qquad \ldots\ldots(18)$$

with the help of (8), (9) and (10). This result does not depend on the assumption of thermodynamic equilibrium.

The fact, shown by equation (15), that F_q is a much simpler function than f_q in thermodynamic equilibrium, suggests that it might also be easier to determine when equilibrium does not obtain. If this should prove to be so, and the proof will be attempted in §4, equation (6) above may well provide the easiest route to calculating f_q. All relevant physical quantities can be expressed in terms of f_q, and therefore in terms of the F_q. The next section will, in fact, be devoted to showing how various important quantities can be calculated directly when the F_q are known.

§3. PROPERTIES OF THE DISTRIBUTION FUNCTIONS

Let $\theta_1(\mathbf{x}, \boldsymbol{\xi})$ be any function of the position and velocity of a molecule; the average value $\langle\theta_1\rangle$ of this function for molecules in the domain considered is given by

$$\langle q\rangle\langle\theta_1\rangle = \int f_1\theta_1 d\tau_1 \qquad \ldots\ldots(19)$$

where $\langle q\rangle$ is given by equation (11). The right-hand side of this equation can be expressed in terms of the F_q by using (6):

$$\begin{aligned}\langle q\rangle\langle\theta_1\rangle &= \sum_{q=0}^{\infty}(q\,!)^{-1}\int F_{q+1}\theta_1 d\tau_{q+1}\\ &= \sum_{q=1}^{\infty}(q\,!)^{-1}\int F_q\left(\sum_{j=1}^{q}\theta_1^{(j)}\right)d\tau_q \qquad \ldots\ldots(20)\end{aligned}$$

where $\theta_1^{(j)}=\theta_1(\mathbf{x}^{(j)}, \boldsymbol{\xi}^{(j)})$, and the symmetry of the function F_q with respect to positions and velocities has been used. Similarly, if $\theta_2(\mathbf{x}^{(1)}, \mathbf{x}^{(2)};\ \boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)})$ depends on the positions and velocities of a pair of molecules, its average value $\langle\theta_2\rangle$ for pairs within the volume V is given by

$$\begin{aligned}\langle q(q-1)\rangle\langle\theta_2\rangle &= \int f_2\theta_2 d\tau_2\\ &= 2\sum_{q=2}^{\infty}(q\,!)^{-1}\int F_q\left(\sum_{j>i}^{q}\theta_2^{(ij)}\right)d\tau_q, \qquad \ldots\ldots(21)\end{aligned}$$

where $\theta_2^{(ij)}=\theta_2(\mathbf{x}^{(i)}, \mathbf{x}^{(j)};\ \boldsymbol{\xi}^{(i)}, \boldsymbol{\xi}^{(j)})$, and $\langle q(q-1)\rangle$ is given by (11) and (12).

As an application of these formulae, the mean energy E of the fluid within V, excluding surface energy, can be expressed in either of the forms

$$E=\tfrac{1}{2}m\int f_1\boldsymbol{\xi}^{(1)2}d\tau_1+\tfrac{1}{2}\int f_2\Phi_2 d\tau_2 \qquad \ldots\ldots(22)$$

or

$$E=\sum_{q=1}^{\infty}(q\,!)^{-1}\int F_q E_q d\tau_q \qquad \ldots\ldots(23)$$

with E_q given by (17). The internal energy is obtained by subtracting the macroscopic energy of flow, which is

$$\tfrac{1}{2}m\int f_1\mathbf{u}^{(1)2}d\tau_1=\tfrac{1}{2}m\sum_{q=1}^{\infty}(q\,!)^{-1}\int F_q\left(\sum_{j=1}^{q}\mathbf{u}^{(j)2}\right)d\tau_q \qquad \ldots\ldots(24)$$

where $\mathbf{u}^{(j)}$ is the velocity of flow at the point $\mathbf{x}^{(j)}$. It is worth noticing that, if

$$\mathbf{v}^{(j)}=\boldsymbol{\xi}^{(j)}-\mathbf{u}^{(j)} \qquad \ldots\ldots(25)$$

is the molecular velocity relative to the fluid motion

$$\int f_1\mathbf{v}^{(1)}d\tau_1=\sum_{q=1}^{\infty}(q\,!)^{-1}\int F_q\left(\sum_{j=1}^{q}\mathbf{v}^{(j)}\right)d\tau_q=0. \qquad \ldots\ldots(26)$$

The mean pressure tensor in V, correct to terms linear in the gradient of $\mathbf{u}$, is given by

$$V\mathbf{p}=m\int f_1\mathbf{v}^{(1)}\mathbf{v}^{(1)}d\tau_1-\tfrac{1}{2}\int f_2\left(\mathbf{x}^{(1)}\frac{\partial\Phi_2}{\partial\mathbf{x}^{(1)}}+\mathbf{x}^{(2)}\frac{\partial\Phi_2}{\partial\mathbf{x}^{(2)}}\right)d\tau_2 \qquad \ldots\ldots(27)$$

or

$$V\mathbf{p}=\sum_{q=1}^{\infty}(q\,!)^{-1}\int F_q\sum_{j=1}^{q}\left(m\mathbf{v}^{(j)}\mathbf{v}^{(j)}-\mathbf{x}^{(j)}\frac{\partial\Phi_q}{\partial\mathbf{x}^{(j)}}\right)d\tau_q. \qquad \ldots\ldots(28)$$

If, then, F_q can be determined to terms linear in the gradient of $\mathbf{u}$, the coefficient of viscosity can be calculated from (28). Similarly the mean thermal flux $\mathbf{q}$ in V is given by

$$V\mathbf{q}=\tfrac{1}{2}m\int f_1\mathbf{v}^{(1)}\mathbf{v}^{(1)2}d\tau_1$$

$$-\tfrac{1}{2}\int f_2\left(\mathbf{x}^{(1)}\mathbf{v}^{(1)}\cdot\frac{\partial\Phi_2}{\partial\mathbf{x}^{(1)}}+\mathbf{x}^{(2)}\mathbf{v}^{(2)}\cdot\frac{\partial\Phi_2}{\partial\mathbf{x}^{(2)}}\right)d\tau_2 \qquad \ldots\ldots(29)$$

if the origin of coordinates is taken at the instantaneous mean centre of V, and this transforms to

$$V\mathbf{q}=\sum_{q=1}^{\infty}(q\,!)^{-1}\int F_q\sum_{j=1}^{q}\left(\tfrac{1}{2}m\mathbf{v}^{(j)}\mathbf{v}^{(j)2}-\mathbf{x}^{(j)}\mathbf{v}^{(j)}\cdot\frac{\partial\Phi_q}{\partial\mathbf{x}^{(j)}}\right)d\tau_q. \qquad \ldots\ldots(30)$$

The coefficient of thermal conduction can be obtained from this formula, if F_q is known to terms linear in the temperature gradient.

Enough has already been done to show that most macroscopic quantities of interest can be calculated, provided the F_q are known. It is not yet clear, however, that the entropy can be obtained in a similar way. The usual definition is

$$S=-k(N!)^{-1}\int f_N\ln f_N d\tau_N \qquad \ldots\ldots(\text{A }1)$$

on the assumption that there are precisely N molecules in the fluid, and the whole fluid is contained in the region of integration. Neither of these assumptions is wanted in the present theory, and the above definition has to be abandoned. It is important, however, that any new definition should retain the important properties of the old one. Thus S should reduce to its known value for thermodynamic equilibrium when the F_q, like f_N, are known. Secondly, the entropy of any non-equilibrium state should be less than that of the equilibrium state with the same mean density and internal energy. Thirdly, the entropy of a composite region in the fluid should be the sum of the entropies of the constituent regions, provided surface effects are neglected. It is also desirable that the total entropy should increase with time, even for systems without external interactions; the formula (A1) does not insure this property.

One might attempt to satisfy these requirements without modifying the old definition in any essential respect. It can be written (see Green 1952, p 73) in the form of an infinite series:

$$S=-k\sum_{q=1}^{\infty}(q\,!)^{-1}\int f_q z_q d\tau_q \qquad \ldots\ldots(\text{A2})$$

where

$$\left.\begin{aligned} z_1&=\ln f_1,\\ z_2&=\ln f_2-\ln f_1^{(1)}-\ln f_1^{(2)},\\ z_3&=\ln f_3-\ln f_2^{(12)}-\ln f_2^{(23)}-\ln f_2^{(31)}\\ &\quad+\ln f_1^{(1)}+\ln f_1^{(2)}+\ln f_1^{(3)}, \end{aligned}\right\} \qquad \ldots\ldots(\text{A3}$$

etc. In this form the assumption that $f_{N+1}=0$ is suppressed and it is no longer possible to prove that S is stationary; no one has yet shown that it increases, either! But if one attempts to transcribe the formula (A 2) in terms of the function F_q, the result is intolerably complicated. So it seems advisable to abandon the old definition entirely, and write instead

$$S=-k\sum_{q=0}^{\infty}(q!)^{-1}\int F_q \ln F_q d\tau_q. \qquad \ldots\ldots(31)$$

This is, like other macroscopic quantities already considered, the sum of a number of terms, each of which is associated with the possibility of finding a given number of molecules in V. Moreover, each of these terms has the suggestive form $-k\Sigma p_r \ln p_r$, where p_r is the probability of a special configuration in phase space.

It must be shown that (31) is exact for thermodynamic equilibrium. To do this, (15) and (13) are used to obtain

$$\left.\begin{aligned} S&=-k\sum_{q=0}^{\infty}(q!)^{-1}\int F_q\beta(q\chi-E_q-pV)d\tau_q\\ &=-T^{-1}(\langle q\rangle\chi-E-pV)\end{aligned}\right\} \qquad \ldots\ldots(32)$$

with the help of (8), (9), (10), (11) and (23). Since $\langle q\rangle\chi$ is the total Gibbs thermodynamic potential in V, a correct thermodynamic relation results. One can show in the usual way that every term of (31) has its maximum value in equilibrium, for specified values of V and E, and also that if $V=V'+V''$ so that $F_q=F_a{}'F_{q-a}{}''$, where a is the number of coordinates in V', then $S=S'+S''$. On these counts the definition (31) is therefore as satisfactory as (A1). If it can be shown further that S naturally increases in non-uniform fluids, there will be sufficient reason to warrant its final adoption.

§ 4. THEORY OF NON-UNIFORM FLUIDS

The next object in view is to formulate and solve an equation, analogous to (1) which is satisfied by F_q, irrespective of the state of the fluid. From (7) it follows that

$$\frac{\partial F_q}{\partial t}=\sum_{j=0}^{\infty}(-1)^j(j!)^{-1}\int^{(j)}\!\!\ldots\int\left\{\frac{\partial f_q}{\partial t}+\sum_{i=1}^{j}\frac{\partial(f_{q+i}\mathbf{u}^{(q+1)})}{\partial\mathbf{x}^{(q+1)}}\right\}d\tau^{(q+1)}\ldots d\tau^{(q+j)}, \qquad \ldots\ldots(33)$$

the second term within the integral arising because the surface S is moving with velocity $\mathbf{u}$ at every point. The time derivative can be eliminated from the first term with the help of (1). Terms involving derivatives with respect to $\boldsymbol{\xi}^{(q+1)},\ldots\boldsymbol{\xi}^{(q+j)}$ then vanish on integration, and some of the remaining terms cancel so one obtains

$$\left.\begin{aligned}\frac{\partial F_q}{\partial t}=\sum_{j=0}^{\infty}(-1)^j(j!)^{-1}\int^{(j)}\!\!\ldots\int&\left\{\left(\frac{\partial}{\partial t}-\frac{d_q}{dt_q}\right)f_{q+j}\right.\\ &\left.-\sum_{i=q+1}^{q+j}\frac{\partial(f_{q+j}\mathbf{v}^{(i)})}{\partial\mathbf{x}^{(i)}}\right\}d\tau^{(q+1)}\ldots d\tau^{(q+j)}.\end{aligned}\right\} \qquad \ldots\ldots(34)$$

Using (7) again to eliminate f_{q+j}, one has finally

$$\frac{d_qF_q}{dt_q}=\int\frac{\partial.(F_{q+1}\mathbf{v}^{(q+1)})}{\partial\mathbf{x}^{(q+1)}}d\tau^{(q+1)}. \qquad \ldots\ldots(35)$$

On using Gauss' theorem, this becomes

$$\frac{d_q F_q}{dt_q} = \int F_{q+1} d\sigma^{(q+1)} \qquad \ldots\ldots(36)$$

where $d\sigma^{(q+1)}$ means the product of $d\boldsymbol{\xi}^{(q+1)}$ and the molecular flux $\mathbf{v}^{(q+1)} . \mathbf{ds}^{(q+1)}$ through the vector surface element $\mathbf{ds}^{(q+1)}$. On account of (26), (36) can also be written

$$\frac{d_q F_q}{dt_q} = \int (F_{q+1} - F_q f_1^{(q+1)}) d\sigma^{(q+1)}. \qquad \ldots\ldots(37)$$

Now it has already been seen, from (18), that

$$F_{q+1} = F_q f_1^{(q+1)} \qquad (d\sigma^{(q+1)} < 0) \qquad \ldots\ldots(38)$$

expresses the statistical postulate that a molecule entering V will have its position and velocity completely uncorrelated with those already in V. In order that this postulate should be rigorously justified, it would be necessary that a molecule outside V could not interact with those inside. It will, in fact, be supposed that the interaction of a molecule with others inside V is 'switched' on or off as it enters or leaves the region. Any error arising from this assumption will obviously affect only surface contributions to volume-dependent quantities.

The factorization indicated by (38) is not possible if $d\sigma^{(q+1)} > 0$, and the postulate so formulated is therefore time-irreversible. It is, in fact the only time-irreversible premise of the theory. Its effect is to exclude from the right-hand side of (37) contributions arising from molecules entering V, and in fact integrations over $d\sigma^{(q+1)}$ *will henceforth be restricted to positive values of this flux element*. Intuitively one should expect that the chain of equations thus obtained, together with

$$f_1 = F_1 + \int F_2 d\tau^{(2)} + \ldots \qquad \ldots\ldots(39)$$

are soluble in principle. It is not the purpose of this paper to provide a rigorous proof that this is so, but it will be shown that the equations enable one to express F_{q+1} exactly in terms of F_q, and some indication will afterwards be given of how one might attempt a practical solution.

To express F_{q+1} in terms of F_q, set

$$\alpha = \int f_1 d\sigma \qquad \ldots\ldots(40)$$

so that α is the total molecular flux out of, or, what is the same thing, into the region considered. Then (37), with $q+1$ replacing q, can be written

$$\frac{d_{q+1} F_{q+1}}{dt_{q+1}} + \alpha F_{q+1} = \int F_{q+2} d\sigma^{(q+2)}. \qquad \ldots\ldots(41)$$

Let $\mathbf{x}_1^{(i)}$ and $\boldsymbol{\xi}_1^{(i)}$ $(i = 1, 2, \ldots q+1)$ be the positions and velocities at time t_1 of a set of $q+1$ molecules, moving under their mutual interactions, if their positions and velocities at time t are $\mathbf{x}^{(i)}$ and $\boldsymbol{\xi}^{(i)}$; they can be calculated in principle by integrating the equation of motion. Further, let $F_{q+1}(t_1)$ and $F_{q+2}(t_1)$ represent the functions F_{q+1} and F_{q+2}, with t_1, $\mathbf{x}_1^{(i)}$ and $\boldsymbol{\xi}_1^{(i)}$ as arguments instead of t, $\mathbf{x}^{(i)}$ and $\boldsymbol{\xi}^{(i)}$. Then it follows from (41) that

$$F_{q+1} = I(t_1, t) F_{q+1}(t_1) + \int_{t_1}^{t} \int I(t', t) F_{q+2}(t') d\sigma^{(q+2)} dt' \qquad \ldots\ldots(42)$$

where $F_{q+2}(t')$ is $F_{q+2}(t_1)$ with t' substituted for t_1, and

$$I(t_1, t) = \exp\left(-\int_{t_1}^{t} \alpha(t')dt'\right). \qquad \ldots\ldots(43)$$

Iteration of this result gives

$$\left.\begin{aligned} F_{q+1} = {} & I(t_1, t)F_{q+1}(t_1) + \int_{t_1}^{t}\int I(t_2, t)F_{q+2}(t_2)d\sigma^{(q+2)}dt' \\ & + \int_{t_1}^{t}\int\int_{t_2}^{t'}\int I(t_3, t)F_{q+3}(t_3)d\sigma^{(q+3)}dt''d\sigma^{(q+2)}dt' + \ldots \end{aligned}\right\} \qquad \ldots\ldots(44)$$

where $F_{q+2}(t_2)$ is derived from $F_{q+2}(t')$ in the same way as $F_{q+1}(t_1)$ is derived from F_{q+1}, and in general $F_{q+j+1}(t_{j+1})$ is derived from $F_{q+j+1}(t^{(j)})$ in the same way. Thus $\mathbf{x}_{j+1}^{(i)}$, $\boldsymbol{\xi}_{j+1}^{(i)}(i=1, 2, \ldots q+j+1)$ and t_{j+1} depend implicitly on $t^{(j)}$, $t^{(j-1)}, \ldots$ and t' as well as $\mathbf{x}^{(i)}$, $\boldsymbol{\xi}^{(i)}$ and t. So far, t_1, t_2, etc. have not been specified. Now t_1 will be chosen to be the last time before t when one of the $\mathbf{x}_1^{(i)}$, $\mathbf{x}_1^{(a)}$ say, was on the boundary of V; then, according to (38),

$$F_{q+1}(t_1) = F_q^{[a]}(t_1)f_1^{(a)}(t_1) \qquad \ldots\ldots(45)$$

where $F_q^{[a]}(t_1)$ is the function F_q of the variables $\mathbf{x}_1^{(i)}$, $\boldsymbol{\xi}_1^{(i)}$ $(i=1, 2, \ldots q+1)$ and t_1, with $\mathbf{x}_1^{(a)}$ and $\boldsymbol{\xi}_1^{(a)}$ omitted. Similarly, t_2 will be chosen to be the next to last time before t' when one of the $\mathbf{x}_2^{(i)}$, $\mathbf{x}_2^{(b)}$ say, was on the boundary; so if t_2' was the last time,

$$F_{q+2}(t_2) = F_q^{[ab]}(t_2)f_1^{(b)}(t_2)f_1^{(a)}(t_2') \qquad \ldots\ldots(46)$$

where $t_2 < t_2' < t'$. In general

$$F_{q+j+1}(t_{j+1}) = F_q^{[ab\ldots z]}(t_{j+1})f_1^{(z)}(t_{j+1})\ldots f_1^{(a)}(t_{j+1}^{(j)}) \qquad \ldots\ldots(47)$$

where $t_{j+1} < t_{j+1}' < \ldots < t_{j+1}^{(j)} < t^{(j)}$. When (45), (46) and (47) have been substituted in (44), the result is the required expression for F_{q+1} in terms of F_q. If this again is substituted into (37), the result is

$$\begin{aligned} \frac{d_qF_q}{dt_q} = \sum_{j=0}^{\infty}\int\int_{t_1}^{t}\cdots\int_{t_j}^{t^{(j-1)}} & I(t_{j+1}, t)[F_q^{[ab\ldots z]}(t_{j+1})f_1^{(z)}(t_{j+1})\ldots f_1^{(a)}(t_{j+1}^{(j)}) \\ & - F_q(t)f_1^{(q+1)}(t)\ldots f_1^{(q+j+1)}(t^{(j)})]d\tau^{(q+j+1)}(t^{(j)})\ldots d\tau^{(q+2)}(t)d\sigma^{(q+1)} \end{aligned} \quad .(48)$$

in which $d\tau^{(q+i+1)}(t^{(i)})$ is short for $d\sigma^{(q+i+1)}dt^{(i)}$. This is a formal generalization of Boltzmann's equation in gas theory, where, since the density is small, F_1 can be approximated by f_1, if V is not too large, and $I(t_1, t)$ is not very different from 1.

Each term on the right-hand side of (48) represents a possible history of how q molecules came to be left in V by the successive departure of additional molecules. The solution of this equation or this set of equations for $q=1$, 2 etc. is, however, obviously beset by great practical difficulties, even presupposing that the techniques developed for Boltzmann's equation (see Chapman and Cowling 1939) could be applied. Firstly, the function f_1 which appears depends, according to (39), on all the F_q. However, it would probably be a good approximation to replace (39) by

$$P_{\langle q\rangle}f_1 = \int^{\langle q-1\rangle}\cdots\int F_{\langle q\rangle}d\tau^{(q)}\ldots d\tau^{(2)} \qquad \ldots\ldots(49)$$

where $\langle q\rangle$ is the most probable number of molecules in V. This is an analogue of the procedure in statistical mechanics of approximating a series such as (39) by its largest term. Another formidable practical difficulty would be to integrate the equations of q interacting molecules, when q is even moderately large. Here again, however, one could use the intuitive fact that the times $t-t_j$ are, on the average, very short, and apply perturbation theory to the problem. The solution of (48), at least for $q=\langle q\rangle$, is thus probably less technically difficult, to a fair degree of approximation, than might at first appear. The remainder of this paper is devoted to a proof that this equation does correctly predict irreversible processes.

§ 5. THE *H*-THEOREM

The object of this section is to show that the entropy, defined by (31), increases in a fluid which is not in thermodynamic equilibrium.

More specifically, it will be shown that

$$\Delta H=\sum_q (q\,!)^{-1}\int_{\Delta t}\int\left(\frac{\partial}{\partial t}+\sum_{i=1}^{q}\boldsymbol{\xi}^{(i)}\cdot\frac{\partial}{\partial\mathbf{x}^{(i)}}\right)(F_q\ln F_q)d\tau_q dt \qquad \ldots\ldots(50)$$

defines an essentially negative quantity ΔH. It will be assumed that F_q does not depend explicitly upon the time, the time dependence arising from the variation of such macroscopic parameters as the local molecular density, temperature and fluid velocity. If ΔH is identified with the negative increase in entropy (divided by Boltzmann's constant k) in the time Δt, it can be shown (for example see Green 1952, pp. 122–125) on the assumption that ordinary thermodynamics can be applied as a first approximation to steady-state irreversible phenomena (discussed in detail by de Groot 1951, pp. 9–12), that

$$-\frac{\Delta H}{\Delta t}=\int\left\{\frac{\lambda}{k}\left(\frac{1}{\beta}\left(\frac{\partial\beta}{\partial\mathbf{x}}\right)\right)^2+2\beta\eta\left(\frac{\partial_s\mathbf{u}}{\partial\mathbf{x}}\right)^2+\beta\eta_0\left(\frac{\partial\,.\,\mathbf{u}}{\partial\mathbf{x}}\right)^2\right\}d\tau, \qquad \ldots(51)$$

where λ is the coefficient of thermal conduction, $\beta=(kT)^{-1}$, and η and η_0 are the coefficients of shearing and volume viscosity. A positive value of $-\Delta H$ therefore implies the existence of viscosity and thermal conduction.

For the purpose of proving the *H*-theorem, a number of identities will be required. The first,

$$\sum_{q=0}^{\infty}(q\,!)^{-1}\int F_{q+1}d\tau_q^{[1]}=\sum_{q=0}^{\infty}(q\,!)^{-1}\int F_q^{[1]}f_1^{(1)}d\tau_q^{[1]} \qquad \ldots\ldots(52)$$

in which $d\tau_q^{[1]}=d\tau^{(2)}\ldots d\tau^{(q+1)}$, follows from the fact that, according to (6) with $q=1$ and (10), each side is equal to $f_1^{(1)}$. The correctness of the second,

$$\begin{aligned}&\sum_{q=0}^{\infty}(q\,!)^{-1}\int\int I(t',t)F_{q+2}(t')d\tau_q^{[12]}d\tau^{(1)}(t)\\ &=\sum_{q=0}^{\infty}(q\,!)^{-1}\int\int I(t',t)F_q^{[1,\,2]}f_1^{(1)}f_1^{(2)}(t')d\tau_q^{[12]}d\tau^{(1)}(t)\end{aligned} \qquad \ldots\ldots(53)$$

where $d\tau_q^{[12]}=d\tau^{(3)}\ldots d\tau^{(q+2)}$, is seen from the fact that each side represents the probability of finding a molecule with an assigned position and velocity at time t', subject to the condition that a molecule will leave the volume V before one enters. The general relation is

$$\begin{aligned}&\sum_{q=0}^{\infty}(q\,!)^{-1}\int\ldots\int I(t^{(j)},t)\,\{F_{q+j+1}(t^{(j)})-F_q^{[1,\,\ldots\,j+1]}f_1^{(1)}\ldots\\ &\qquad f_1^{(j)}(t^{(j-1)})d\tau_q^{[1,\,\ldots\,j+1]}d\tau^{(1)}(t)\ldots d\tau^{(j)}(t^{(j-1)})=0.\end{aligned} \qquad \ldots\ldots(54)$$

Now, returning to (50), the integrand of this equation can be replaced by $(d_q/dt_q)(F_q \ln F_q)$, since derivatives with respect to the velocities will disappear on integration. On account of (10), this may in turn be replaced by the term $(\ln F_q)(d_q F_q/dt_q)$. Substituting from (37) for $d_q F_q/dt_q$, one has, therefore,

$$\Delta H = \sum_q (q!)^{-1} \int\int_{\Delta t} \ln F_q(F_{q+1} - F_q f_1^{(q+1)}) d\tau^{(q+1)}(t) d\tau_q. \quad \ldots\ldots(55)$$

To the right-hand side of this equation one now adds the expression

$$\sum_q (q!)^{-1} \int\int_{\Delta t} \ln f_1^{(q+1)}(F_{q+1} - F_q f_1^{(q+1)}) d\tau^{(q+1)}(t) d\tau_q$$

which vanishes, according to the identity (52). One next substitutes (42) for F_{q+1}, and adds a term

$$\sum_q (q!)^{-1} \int\int_{\Delta t} \int_{t_1}^{t} I(t', t) \ln f_1^{(q+2)}(t') \{F_{q+2}(t') - F_q f_1^{(q+1)} f_1^{(q+2)}(t')\} \\ \times d\tau^{(q+2)}(t') d\tau^{(q+1)}(t) d\tau_q$$

which also vanishes, according to (53). Proceeding in this way, one obtains finally

$$\Delta H = \sum_q (q!)^{-1} \sum_j \Delta K_{q,j} \quad \ldots\ldots(56)$$

where

$$\Delta K_{q,j} = \int\int_{\Delta t} \cdots \int_{t_j}^{t(j-1)} I(t_{j+1}, t) \ln \{F_q f_1^{(q+1)} \ldots f_1^{(q+j+1)}(t^{(j)})\} \\ \times [\ldots] d\tau^{(q+j+1)}(t^{(j)}) \ldots d\tau^{(q+1)}(t) d\tau_q \quad \ldots\ldots(57)$$

where [...] indicates the integrand of (48).

To prove the *H*-theorem, one has now only to show that the above integral remains unchanged if one interchanges the arguments of F_q and $F_q^{[ab\ldots z]}(t_{j+1})$ and of the $f_1^{(q+i+1)}(t^{(i)})$ and $f_1^{(p)}(t_{j+1}^{(i)})(i = 0, 1, \ldots j)$ simultaneously. This can be done by a succession of three transformations:

(1) Let the variables $\mathbf{x}_q$ (i.e. $\mathbf{x}^{(1)} \ldots \mathbf{x}^{(q)}$), $\boldsymbol{\xi}_q$, $\mathbf{x}^{(q+i+1)}(t^{(i)})$ and $\boldsymbol{\xi}^{(q+i+1)}(t^{(i)})$ in (57) be transformed by elementary substitution to

$$-\mathbf{x}_q^{[ab\ldots z]}(-t_{j+1}), \quad \boldsymbol{\xi}_q^{[ab\ldots z]}(-t_{j+1}), \quad -\mathbf{x}^{(p)}(-t_{j+1}^{(i)}) \quad \text{and} \quad \boldsymbol{\xi}^{(p)}(-t_{j+1}^{(i)}),$$

where p is by definition the $(i+1)$th of the superfixes $z, \ldots, b, a$; and let the limits of integration in space be similarly transformed so that the value of the integral (57) is unchanged. Since the motion of any set of mutually interacting molecules, reflected through the mean centre of V, is reversible, the transformed values of $\mathbf{x}_q^{[ab\ldots z]}(t_{j+1})$, $\boldsymbol{\xi}_q^{[ab\ldots z]}(t_{j+1})$, $\mathbf{x}^{(p)}(t_{j+1}^{(i)})$ and $\boldsymbol{\xi}^{(p)}(t_{j+1}^{(i)})$ will be $-\mathbf{x}_q(-t)$, $\boldsymbol{\xi}_q(-t)$, $-\mathbf{x}^{(q+i+1)}(-t^{(i)})$ and $\boldsymbol{\xi}^{(q+i+1)}(-t^{(i)})$.

(2) The independent variables of integration are now changed from $-\mathbf{x}_q^{[ab\ldots z]}(-t_{j+1})$, $\boldsymbol{\xi}_q^{[ab\ldots z]}(-t_{j+1})$, $-\mathbf{x}^{(p)}(-t_{j+1}^{(i)})$ and $\boldsymbol{\xi}^{(p)}(-t_{j+1}^{(i)})$ to $-\mathbf{x}_q(-t)$, $\boldsymbol{\xi}_q(-t)$, $-\mathbf{x}^{(q+i+1)}(-t^{(i)})$ and $\boldsymbol{\xi}^{(q+i+1)}(-t^{(i)})$. This affects only the limits and variables at integration, since by Liouville's theorem any volume element of phase space is unchanged by motion in accordance with mechanical laws.

(3) The sign of all coordinate variables and t and $t^{(i)}$ is reversed, and the limits of integration altered to return to their original values. Since F_q and f_1 do not depend explicitly on the time, the final result is

$$\Delta K_{q,j} = -\int\int_{\Delta t} \ldots \int_{t_j}^{t(j-1)} I(t_{j+1}, t) \ln \{F_q^{[ab\ldots z]}(t_{j+1}) f_1^{(z)}(t_{j+1}) \ldots f_1^{(a)}(t_{j+1}^{(j)})\}[\ldots] d\tau^{(q+j+1)}(t^{(j)}) \ldots d\tau^{(q+1)}(t) d\tau_q. \qquad \ldots\ldots(58)$$

Taking half the sum of (57) and (58), one obtains a negative definite expression for $\Delta K_{q,j}$. This completes the proof of the H-theorem.

The physical meaning of this result is that every cycle beginning and ending with q molecules in V, and intermediately allowing the entry of any number of molecules and the departure of a similar number, makes its own contribution to the increase of entropy. The original H-theorem is the special case with $q=1$, where only one molecule is allowed to enter and depart from the region of interaction of the original molecule.

REFERENCES

BORN, M., and GREEN, H. S., 1946, *Proc. Roy. Soc.* A, **188**, 10 ; 1947, *Ibid.*, **190**, 455.

CHAPMAN, S., and COWLING, T. G., 1939, *The Mathematical Theory of Non-uniform Gases* (Cambridge : University Press).

COX, R. T., 1950, *Rev. Mod. Phys.*, **22**, 238.

DAVYDOV, B., 1947, *J. Phys., U.S.S.R.*, **11**, 33.

GREEN, H. S., 1952, *Molecular Theory of Fluids* (Amsterdam : North-Holland Publishing Company).

DE GROOT, S. R., 1951, *Thermodynamics of Irreversible Processes* (Amsterdam : North-Holland Publishing Company).

IRVING, J. H., and KIRKWOOD, J. G., 1950, *J. Chem. Phys.*, **18**, 817.

KIRKWOOD, J. G., 1946, *J. Chem. Phys.*, **14**, 180.

KIRKWOOD, J. G., BUFF, F. P., and GREEN, M. S., 1949, *J. Chem. Phys.*, **17**, 988.

KIRKWOOD, J. G., and SALSBURG, Z. W., 1953, *Disc. Faraday Soc.*, No. 15, 28.

KLEIN, G., and PRIGOGINE, I., 1953, *Physica*, **19**, 74, 89, 1053.

MAYER, J. E., 1942, *J. Chem. Phys.*, **10**, 629.

TOLMAN, R. C., 1938, *The Principles of Statistical Mechanics* (Oxford : University Press).

ZWANZIG, R. Q., KIRKWOOD, J. G., OPPENHEIM, I., and ALDER, B. J., 1954, *J. Chem. Phys.*, **22**, 783.

Nuclear Magnetic Resonance Absorption in Polyisobutylene

By J. G. POWLES†

Physics Department, King's College, University of Durham, Newcastle-upon-Tyne

MS. received 27th September 1955 *and in amended form 2nd November* 1955

Abstract. The proton magnetic resonance absorption in polyisobutylene has been measured over the temperature range −196°C to 90°C. The unusually large width of the absorption line at the lowest temperatures is interpreted as largely due to the well known close approach of the methyl groups on alternate carbon atoms in the chain. The usual reduction in line width associated with reorientation of methyl groups about their C_3 axis is notably absent and supports the contention that the methyl groups are severely interlocked. A reduction in the second moment of the absorption line which sets in over a range of temperature near −10°C is associated with chain motion. A satisfactory correlation has been obtained between this observation and measurements of mechanical properties. Experiments on PIB swelled with benzene substantiate the interpretation given above and suggest that valuable information as to the nature of the swelling process of rubber-like polymers may be obtained by means of nuclear resonance studies.

§ 1. Introduction

The basic nuclear magnetic resonance experiment consists in placing a sample containing nuclei having a magnetic moment (and consequently a non-zero spin) in a magnetic field usually of several thousand oersteds. This causes the nuclei to precess about the field direction at the Larmor frequency which usually falls in the radio-frequency region. The angle between the nuclear axis and the magnetic field, and therefore the energy of the nuclear system, can be changed by means of another field, at right angles to the large field, which rotates at the Larmor frequency. If this field is provided by a coil the energy absorbed by the agency of the nuclei is detected as a small change in the losses in the coil near the Larmor frequency by appropriate electronic techniques. Various modifications of this basic experiment are used and the one used in the present investigation is described in the next section.

Nuclear magnetic resonance absorption measurements are particularly valuable in solid state physics since the absorption is dependent, among other things, on the relative positions of the nuclei, their arrangement relative to the large magnetic field, and on the motion of the nuclei. Information on the relative positions of protons in a solid is particularly valuable since this is provided by x-ray analysis only with the most refined techniques. In this paper we discuss the relative positions of protons in the long chain hydrocarbon polymer polyisobutylene, for which establishing even the positions of the carbon nuclei has presented some difficulty. The nuclear resonance experiments give less detailed information than would X-ray analysis if applicable to protons, but the results can be used to test various alternative structures suggested by X-ray work. In the case of polyisobutylene the nuclear resonance absorption indicated quite directly the unusually close approach of certain protons in the molecule, which could only be surmised from the X-ray analysis.

---

†Now at Physics Department, Queen Mary College, University of London, Mile End Road, London, E.1

The effect of motion of the nuclei, if sufficiently rapid, is to narrow the absorption line and some indication is obtained of the type, extent and rate of motion involved. In the present case a large effect on the nuclear absorption is caused by motion of the polymer chains. Chain motion also largely determines the mechanical properties in the rubbery state of the polymer. The present experiments, on magnetic properties at 16 Mc/s, confirm the observation of a mechanical absorption in polyisobutylene at 200 c/s. The nuclear magnetic resonance absorption may also be quite sensitive to motions or displacements of nuclei which hardly affect many other properties of the material and this is illustrated by the very considerable changes in absorption observed in polyisobutylene at low temperatures.

§ 2. EXPERIMENTAL ARRANGEMENT AND RESULTS

The large magnetic field was provided by a semi-permanent magnet having a field of 3782 oersteds with an inhomogeneity of about 0·2 oersted in a region of about one cm cube at the centre of the gap. When the resonance line was narrow, use of a small specimen permitted the measurement of line widths down to 0·1 oersted. This homogeneity was obtained by using small coils carrying adjustable currents stuck to the pole faces such that the field produced compensated the inhomogeneity of the permanent magnet field. The magnet was provided with one set of coils by means of which the field could be varied linearly with time over any range up to 50 oersteds at any desired rate and another set by which the field could be modulated at 175 c/s up to 4 oersteds peak to peak. The nuclear resonance signal, at about 16 Mc/s for protons, due to a sample usually about 8 mm cube was detected either by means of an autodyne oscillator circuit (Hopkins 1949) or by means of a bridged T arrangement (Tuttle 1940) with cascode pre-amplifier (Wallman *et al.* 1948) and receiver. In each case the 175 c/s signal was amplified and passed to a synchronous detector (Dicke 1950, Cox 1953) and the resulting direct output fed to a recording potentiometer. The time constant of the detector could be varied between 1¼ and 10 seconds as required by the signal to noise ratio and the rate at which the field was varied.

The autodyne circuit, and the bridged T with suitable unbalance, indicate the energy absorption due to the resonant nuclei. With a modulation amplitude much smaller than the line width, the output at the recorder is proportional to the derivative of the absorption coefficient (χ'') with respect to the large magnetic field. The results are therefore in the form of a plot of derivative of absorption coefficient against field. The samples were held in a cryostat in a Dewar vessel with a 'tail' placed between the poles of the magnet (Gutowsky *et al.* 1953). Any temperature in the range -196°c to 120°c could be obtained and held to about 1°c by using either liquid nitrogen or solid carbon dioxide as coolants, and by adjusting the current in a heater placed near the sample holder.

The majority of the measurements were made with the autodyne oscillator since this arrangement it easiest to operate as a result of its simplicity and its relative insensitivity to vibration and temperature drift. However, because of the difficulty in varying the radio-frequency field applied to the sample to any great extent it is possible that saturation of the nuclear spin system may occur if T_1, the spin–lattice relaxation time, is long enough. This did in fact occur in polyisobutylene (PIB) at temperatures below $-180°$c. All the measurements were checked with the bridged T arrangement with a wide range of radio-frequency fields including that given by the autodyne circuit. The results given are therefore free from line broadening due to saturation. This point is stressed because unusually broad lines were observed in PIB at low temperatures and rather important conclusions rest on this observation, as discussed later.

The second moment $\Delta H_2{}^2$ of the absorption line is of great importance in the interpretation of the results (Van Vleck 1948) and is defined as follows

$$\Delta H_2{}^2 = \frac{\int_{\text{line}} \chi''(H)(H-H_0)^2 dH}{\int_{\text{line}} \chi''(H) dH} \qquad \ldots\ldots(1)$$

where H_0 is the magnetic field at the centre of the line and H is the field. Since we observe $d\chi''(H)/dH$ which we call $\phi(H)$, it is convenient to use the equivalent formula

$$\Delta H_2{}^2 = \frac{1}{3}\,\frac{\int_{\text{line}} \phi(H)(H-H_0)^3 dH}{\int_{\text{line}} \phi(H)(H-H_0) dH}\,. \qquad \ldots\ldots(2)$$

The second moment must be calculated numerically from equation (2) and this becomes extremely tedious. A graphical method based on the fact that equation (2) may be written

$$\Delta H_2{}^2 = \frac{1}{6}\,\frac{\int_{\text{line}} \phi(H) d(H-H_0)^4}{\int_{\text{line}} \phi(H) d(H-H_0)^2} \qquad \ldots\ldots(3)$$

has been developed which is very satisfactory and is reported elsewhere (Powles 1956).

The second moment values were corrected for modulation broadening (Andrew 1953) although in no case did this amount to more than 1% of the second moment.

The accuracy of measurement of $\Delta H_2{}^2$ is independent of its magnitude except in so far as the broad lines are weaker, since the experimental arrangement can be adjusted to accommodate any line width down to a limiting value of about 0·1 oersted determined by the inhomogeneity of the magnet. The error in the second moment values should not be more than $\pm 5\%$.

The samples used (see Acknowledgments) were part of a special batch of polyisobutylene distributed by R S. Marvin of the National Bureau of Standards and many properties of this material have been measured (Marvin 1954). The molecular weight of the sample is $1{\cdot}3 \times 10^6$. It is particularly valuable that the mechanical properties have been extensively studied (Thomas and Robinson 1955).

Figure 1 (*a*) shows the line width δH in oersteds defined as the distance between maxima of $\phi(H)$, as a function of temperature. Where two values are given, $\phi(H)$ has two pairs of peaks. Results are shown for the sample of PIB and for a sample swelled with benzene. The extent of swelling is expressed in terms of monomers of the PIB molecule per molecule of the swelling agent, which is an appropriate unit for the present investigation.

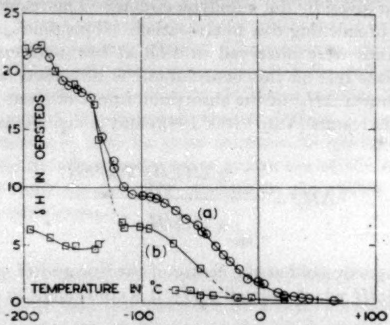


Figure 1. Line width in oersteds against temperature in °c for (*a*) pure polyisobutylene, (*b*) polyisobutylene swelled with benzene, 4·88 monomers of PIB per molecule of benzene.

Figure 2 (*a*) shows the corresponding variation of second moment values, ΔH_2^2, with temperature. These are plotted to the scale of ΔH_2 to illustrate better the wide variation in ΔH_2^2 with temperature. Second moments below about 0·1 oersted², which occur at the higher temperatures, were not measurable because of the field inhomogeneity.

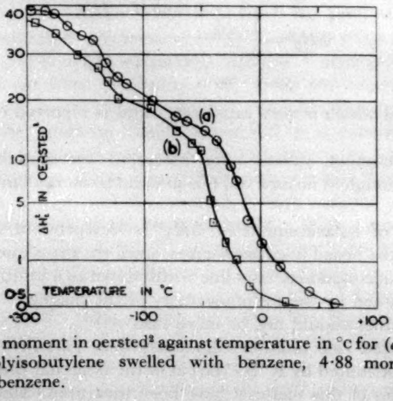


Figure 2. Second moment in oersted² against temperature in °c for (*a*) pure polyisobutylene, (*b*) polyisobutylene swelled with benzene, 4·88 monomers of PIB per molecule of benzene.

Figure 3 shows a plot of the ratio $\delta H/2\Delta H_2$ as a function of temperature and illustrates the change in shape of the absorption curve with temperature. $\delta H/2\Delta H_2$ is unity for a gaussian shaped absorption line.

For pure PIB the second moment is extremely large at liquid nitrogen temperatures. It falls gradually with rising temperature except for a rather sudden but small fall at about −135°. Above −30°C a more rapid fall occurs over a temperature range centred at about −10°C and continues until the line is too narrow to measure with the present apparatus. In the swelled PIB the behaviour is very similar except that the large fall in width occurs at a lower temperature. Each of these observations will be discussed in turn.

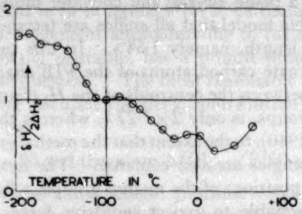


Figure 3. Shape factor $\delta H/2\Delta H_2$ plotted against temperature for pure polyisobutylene.

§ 3. Discussion

3.1. *The Low Temperature Region*

It seems likely that the second moment has reached the rigid lattice value at −196°C. The value attained in PIB is 41 oersted². That this is an unusually high value may be seen by calculating the expected second moment. Remembering that the monomer of PIB is $(CH_3)_2C=CH_2$, we calculate the principal contribution to the second moment using the appropriate form of Van Vleck's formula (1948), eqn (4)

$$\Delta H_2{}^2 = \frac{18}{5}\frac{\mu^2}{N}\sum_{i>j} R_{ij}{}^{-6} \qquad \ldots\ldots(4)$$

where μ is the magnetic moment of the proton in nuclear magnetons, R_{ij} is the separation of the ith and jth protons and N is the number of protons considered to interact.

For interactions of protons within methyl groups and within methylene groups, assuming the C–H distances are 1·09 Å and all angles are tetrahedral, it is found that $\Delta H_2{}^2(CH_3) = 22{\cdot}5$ oersted² and $\Delta H_2{}^2(CH_2) = 11{\cdot}2$ oersted².

Hence $\Delta H_2{}^2(\mathrm{PIB}) = 6/8 \times 22{\cdot}5 + 2/8 \times 11{\cdot}2 = 19{\cdot}7$ oersted². This quantity may be called the internal second moment of PIB. We have also to add to this figure contributions to the second moment due to interactions between protons so far neglected. Since, in general, such proton pairs are separated by at least the van der Waals distance, and the second moment depends on R^{-6}, where R is the separation, this contribution is usually not very large. Nor is the external second moment very important unless the internal second moment is small. If we assume that each proton has two neighbouring protons at the van der Waals distance (2·4 Å) (Pauling 1940) we find a contribution to the second moment of 3·7 oersted². It is found (Powles and Gutowsky 1953) for solid t-butyl chloride, which also has a high concentration of protons, that the experimental value of the external second moment is 5 oersted² so that it is reasonable to expect a contribution of this order of magnitude in PIB in the absence of any unusual effects, although the exact value must depend on the actual arrangement of the protons. We may therefore take the expected rigid lattice second moment of PIB to be about 24 oersted^2. This leaves 17 oersted^2 unaccounted for.

We believe that this discrepancy is largely explained by the well-known, very close approach of the methyl groups which must occur in the PIB molecule (Flory 1937, Bunn 1942, Evans and Polanyi 1943). In other words it is a result of unusually large steric hindrance in the molecule. In order to estimate this special contribution to the second moment we first take the main carbon chain of the molecule to be a plane zig-zag and consider more realistic models later. We shall assume for this model that all angles are tetrahedral and that the C–C bonds are the normal length, namely 1·54 Å. In this case it is found that the methyl groups on alternate carbon atoms of the PIB chain are close together in lines and the distance between the centroids of the H_3 triangles, and of the carbon nuclei of the methyl groups, is only 2 × 1·27 Å, whereas the radius of the methyl group is 2·0 Å (Pauling 1940), to the extent that the methyl group may be considered spherical. The H_3 triangles are also coplanar. The next problem is to decide what are the azimuthal positions of the methyl groups. Because of the very close approach it seems reasonable to expect repulsive forces to be the controlling factor and to assume that the interactions are additive. Two arrangements of adjacent methyl groups which merit consideration are indicated as (*a*) and (*b*) in figure (4). Considering only the shortest distances we have for the proton–proton separations,

$$(a)\quad R_{22'}=R_{33'}=1{\cdot}52\ \text{Å},\qquad R_{23'}=R_{32'}=2{\cdot}37\ \text{Å}.$$

$$(b)\quad R_{31'}=R_{21'}=1{\cdot}33\ \text{Å}.$$

The van der Waals radius of hydrogen is 1·2 Å (Pauling 1940) (i.e. $R_{ij}=2{\cdot}4$ Å), so that if the methyl groups were placed as supposed, very considerable repulsive forces would be involved. Because of the very rapid variation of repulsive force with distance, it is evident that the arrangement (*a*) has the lower energy and is a position of stable equilibrium. A more striking illustration is provided if we assume that the energy of interaction of two protons is

$$E_{ij}=380E(95R_{ij}^{-12}-R_{ij}^{-6}) \qquad \ldots\ldots(5)$$

which corresponds to a binding energy of E at an equilibrium distance of 2·4 Å. It may be shown that the energies E_a and E_b, corresponding to positions (*a*) and (*b*), referred to the minimum of energy, are such that $E_b/E_a=11$. Thus it seems reasonable, in the following discussion, to choose the arrangement which gives the maximum value for the shortest inter-proton distances.

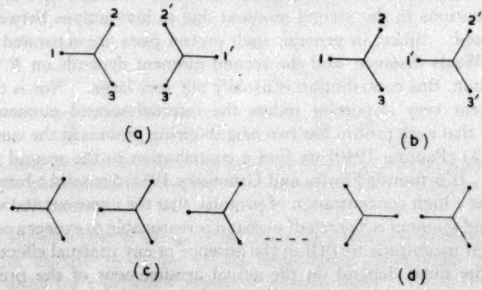


Figure 4. Illustrating various arrangements of methyl groups referred to in the text.

For future reference we note that the 'steric' contributions to the second moment due to these arrangements of a pair of methyl groups are

for (*b*) $\Delta H_2^2(\mathrm{CH_3}) = 45$ and therefore $\Delta H_2^2(\mathrm{PIB}) = 33{\cdot}8$ oersted^2

for (*a*) $\Delta H_2^2(\mathrm{CH_3}) = 20{\cdot}6$ and therefore $\Delta H_2^2(\mathrm{PIB}) = 15{\cdot}5$ oersted^2.

However, in the linear PIB molecule the methyl groups are close together in lines and it is evident by an extension of the argument given above that the lowest energy arrangement is as in figure 4 (*c*) since the occurrence of (*a*) more than once requires an arrangement approximating to (*d*) at some point and this configuration, besides being unstable, has a proton–proton separation of only 0·47 Å which is out of the question. The most reasonable arrangement therefore is as in figure 4 (*c*) and we calculate the steric contribution to the second moment for the infinite chain to be

$$\Delta H_{2\,\mathrm{steric}}^2(\mathrm{CH_3}) = 90 \text{ and therefore } \Delta H_{2\,\mathrm{steric}}^2(\mathrm{PIB}) = 67{\cdot}5 \text{ oersted}^2.$$

We now recall that the extra second moment required by experiment is 17 oersted^2 so that the nuclear resonance data confirm that the straight chain molecule is not acceptable. This, of course, agrees with the considerations of steric hindrance which are discussed in more detail below. A more precise consideration of the methyl group orientations, involving minimization of the free energy and allowing distortion of the chain angles, would make the argument less powerful but would no doubt leave the straight chain as a very unlikely configuration.

The PIB molecules in the polymer are not straight chains and the chain C–C–C angles may not be tetrahedral as a consequence of the steric hindrance. In the polymer the chain is presumably kinked in a substantially random way although certain configurations, particularly straight sections, will be rare because of the large methyl group steric hindrance. Because of this it is not possible to make an accurate estimate of the second moment of the polymer without rather detailed information as to the chain configurations. It is known, however, that PIB crystallizes on stretching (Brill and Halle 1938) and it seems worth while to calculate the second moment for the molecular arrangement then taken up since this might well be a rather favoured arrangement even in the amorphous polymer.

The exact structure of crystalline PIB is not yet decided (Fuller *et al.* 1940, Bunn 1947, Liquori 1955), but there is little doubt that the main carbon chain is approximately helical in form and repeats every eight monomer units. The most likely form is one in which there are five turns of the helix in the repeat distance (S_5) (Bunn 1947, Liquori 1955). A form with seven turns (S_7) is apparently less favoured by the x-ray evidence (Liquori 1955). In S_5 and S_7 it is assumed that there is a uniform helix so that the C–C–C angles in the chain are all equal but not tetrahedral (see table). A rather reasonable supposition is that the repulsion between the methyl groups opens up the C–$\mathrm{CH_2}$–C angle and a model (S_H) having this angle 126° has been proposed (see Acknowledgments). In all these models the methyl groups are close together in pairs only and do not form chains as in the linear molecule.

We have calculated the second moment for these three models. This is less simple than for the straight chain molecule since the H_3 triangles are no longer coplanar, and the difficulty in deciding the angular positions is greater. Moreover some of the methyl and methylene groups are unusually close. To avoid tedious calculations models to a scale of 1 Å = 10 cm were constructed for the close pairs of methyl groups for the three structures, S_5, S_7 and S_H, in which the 'methyl groups' could be rotated about their C_3 axis. The interproton separations of the methyl groups were studied as a function of orientation and the equilibrium configuration chosen by the criterion already discussed, that the shortest interproton distances be given the maximum value. In this way the most likely orientation was chosen and the second moment calculated using equation (4). Although the H_3 triangles are no longer coplanar it is found that the two most likely configurations may again be designated by (*a*) and (*b*) of figure (4), with (*a*) again having much the lower energy.

In the case of the model S_H the rather large value of the chain C–CH_2–C angle makes it necessary to reconsider the internal second moment value for the methylene group. Opening the CCH_2C angle to this extent probably closes the HCH angle from the tetrahedral value (109·5°) to 98·6° (Coulson 1948, Torkington 1951) and may alter the CH distance. Assuming that the CH distance remains unchanged at 1·09 Å we calculate that $\Delta H_2^2(CH_2) = 18{\cdot}4$ and therefore contributes 4·6 oersted^2 to $\Delta H_2^2(PIB)$. The internal second moment is thereby larger by 1·8 oersted^2 than that previously given. This contribution is included in the result (*a*) in column 4 of the table.

Since the model S_H is a very promising one from the x-ray evidence it seems worth while in this case to make a more accurate estimate of the contribution to the second moment other than that due to internal and steric interaction, which has so far been assumed to be about 5·0 oersted^2. It is also desirable because there are several other inter-proton distances in the S_H helix which are less than the van der Waals distance and the effect on the second moment of the greater separation of the methyl groups in this model as compared with S_5 and S_7 is to some extent compensated. Again using the carbon positions provided by Dr. Holmes and with the closure of the HCH angles discussed above, we find that the total second moment for an isolated PIB helix is 38·3 oersted^2. The contribution to the second moment due to interaction between helices should be quite small and this is supported by the experiments on swelled PIB discussed in a later section. It seems reasonable therefore to add at most 2 oersted^2 giving finally a theoretical value of second moment for crystalline PIB of 40·3 oersted^2 which is quoted in column 4 of the table as the result (*b*) for S_H. The result (*b*) is noticeably higher than (*a*) mainly because of the close approach of certain methyl and methylene group protons which a detailed study of S_H reveals. I am grateful to Dr. Holmes for drawing my attention to this point.

Model	(1)	(2)	(3)	(4)
Straight chain	109½	2·54	1·33 (1·52 for (*a*))	92·2
S_7	109	2·37	1·61	39·2
S_5	114	2·31	1·49	45·3
S_H	126	2·79	1·81	(*a*) 33·6 (*b*) 40·6

(1) C–CH_2–C angles in chain ; (2) Distance between centroids of H_3 triangles (Å) ; (3) Shortest inter-proton distances (Å) ; (4) Calculated value of $\Delta H_2^2(PIB)$ (oersted^2).

Bearing in mind that the experimental second moment is 41 oersted^2 with a probable error of some 5% and disregarding, for the moment, the difference between the amorphous and the crystalline polymer, the results in column (4) of the table show that the nuclear resonance experiments are not in violent disagreement with any of the models S_5, S_7 or S_H, and, if anything, they favour the S_H model for PIB. The correct comparison would of course be with the second moment for the crystalline polymer but we have not as yet been successful in preparing a suitable specimen.

It is interesting to note that a consideration of distance between H_3 centroids alone (column (2) of the table) would suggest on steric grounds that S_5 and S_7 are unfavourable alternatives to the straight chain. However, the shortest inter-proton separations are much the same (column (3) of the table) or even greater for S_5, at least for the (*a*) arrangement. The telling argument against the straight chain model for PIB, both on steric and on nuclear magnetic resonance grounds, is that it involves close methyl groups in lines.

We conclude that the nuclear magnetic resonance absorption indicates a very close approach of methyl groups and to a lesser extent of methyl and methylene groups in the amorphous polymer to an extent which is comparable with that found for the molecule in the crystalline form. This suggests that the molecules in the amorphous material may not be very different in form from the helices in the crystal. A departure from the regular arrangement of spirals such as is indicated by the description of the material as 'glassy' in this temperature region and a consideration of the methyl group separation for various chain configurations makes it reasonable to expect a second moment value in the amorphous material not very different from that calculated for the crystalline form, in agreement with the measurements reported here.

3.2. *Variation of Absorption with Temperature*

It is well known (Bloembergen *et al.* 1948) that the contribution to the second moment due to a pair of nuclei falls when the distance between them and/or the angle of the line joining them varies at a rate faster than the line width, δH, expressed in frequency, $\delta\nu$, where $\delta\nu=(\mu/I\hbar)\delta H$.

In figure 2(*a*) the second moment of PIB falls steadily with increasing temperature to a value of 13 oersted^2 at $-30°$c apart from a comparatively sudden fall of about 8 oersted^2 near $-135°$c. In many materials containing methyl groups there is a line narrowing over a comparatively small range of temperature, usually below $-130°$c, which has been interpreted as due to reorientation of the methyl groups about their C_3 axis (Powles and Gutowsky 1953). For PIB this would require the second moment to fall to about 8 oersted^2 since the internal second moment of the methyl protons would fall to one quarter of its rigid lattice value and the steric broadening would undoubtedly be reduced by a large factor. The contribution of the methylene protons would be practically unchanged. The steady fall up to $-30°$c cannot therefore be due to methyl group C_3 reorientation and this accords well with our assumption of severe interlocking of methyl groups. In other words the methyl groups are unable to reorient about the C_3 axis unless the chain segments move appreciably. The narrowing up to $-30°$c may therefore be due to the gradually increasing rate and extent of motion of the molecular chains which allows a certain amount of motion and possibly even some C_3 reorientation of the methyl groups but leaves them essentially restricted. In this case part of the fall in second moment is also due to motion of the methylene groups.

A tentative explanation of the sudden fall of 8 oersted^2 at $-135°$c is that it is due to C_3 reorientation of a small proportion of methyl groups which are not interlocked but are only restrained by normal potential barriers. As already discussed such methyl groups should be comparatively few, and the fall in second moment is consistent with there being about 15% of such methyl groups. On the other hand, the rather important change in the shape of the line (figure 1) together with the comparatively small change in second moment makes it likely that a larger proportion of the protons is involved and that the change is due to a rearrangement of the protons rather than an increase in their rate of motion. In other words there may be a hitherto unrecognized phase transition in this temperature region. There is in fact a small anomaly in the mechanical absorption between −130°c and −150°c (Thomas and Robinson 1955).

Above −30°c the second moment falls rapidly with increasing temperature until it becomes too narrow to observe with the present equipment. Narrowing to this extent must imply a considerable freedom of motion of the chains and the polymer becomes more liquid-like in its resonance absorption. This line width transition is evidently connected with the glass to rubber transition. If we assume that the motion involved may be described by a single correlation time τ_c, or a single characteristic frequency ν_c, and that it is thermally activated the variation of the width with temperature may be used to deduce an activation energy for the process (Gutowsky and Pake 1950). In this case, because of the considerable variation of the line shape (figure 3) over the transition region we prefer to analyse the second moment curve and use the formula (Powles and Gutowsky 1955)

$$\frac{\Delta H_2^2}{\Delta H_{2\,\mathrm{max}}^2} = \frac{2}{\pi} \tan^{-1} \frac{\delta \nu}{\nu_c} . \qquad \ldots\ldots(6)$$

Hence, assuming $\nu_c = \nu_0 \exp(-\Delta E/\boldsymbol{k}T)$ we plot $\ln \tan[\frac{1}{2}\pi \Delta H_2^2/\Delta H_{2\,\mathrm{max}}^2]$ against $1/T$ in figure 5(*a*). The straight line corresponds to $\Delta E = 0{\cdot}44$ ev or 10·2 k cal mole⁻¹ and $\nu_0 = 1{\cdot}9 \times 10^{13}$ c/s. This value of ν_0 corresponds to an entropy of activation (Glasstone *et al.* 1941) $\Delta S = 6{\cdot}1$ e.u.

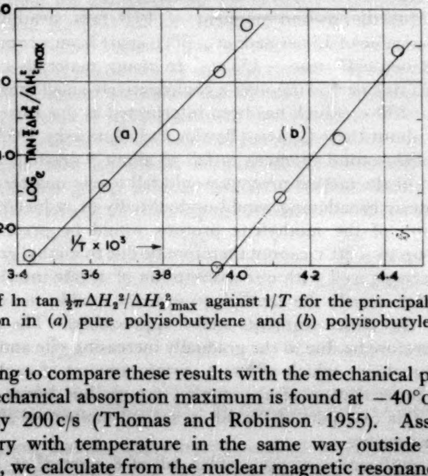


Figure 5. Plot of ln tan $\frac{1}{2}\pi\Delta H_2^2/\Delta H_{2\,\mathrm{max}}^2$ against $1/T$ for the principal line width transition region in (*a*) pure polyisobutylene and (*b*) polyisobutylene swelled with benzene.

It is interesting to compare these results with the mechanical properties of the material. A mechanical absorption maximum is found at −40°c at a frequency of approximately 200 c/s (Thomas and Robinson 1955). Assuming that ν_c continues to vary with temperature in the same way outside the line width transition region, we calculate from the nuclear magnetic resonance experiments, $\nu_c = 187$ c/s at −40°c. The agreement between this figure and the mechanical absorption frequency is probably too good; but it does seem to indicate that the fundamental process involved in the two phenomena is the same or very similar. It will be of great interest to see if the mechanical absorption has the same activation energy. The positive activation entropy is also reasonable since this may be interpreted to mean that a considerable number of chain segments is involved in the activation process. If this is so we may well expect a correlation between the mechanical absorption frequency and the nuclear resonance absorption line width transition. The interpretation of the large positive entropies of activation found for the dielectric relaxation process in rubbers composed of molecules bearing polar groups is very similar (Kauzmann 1942).

The nuclear resonance experiments therefore give a direct indication that considerable motion of the molecular chain is involved in the rubbery region and this is in accord with the rubber-like properties of the material.

In a material as complicated as PIB it is not possible to give any detailed interpretation of the changes in line shape shown in figure 3. The fact that $\delta H/2\Delta H_2$ is not unity at -196°c indicates that the protons are not randomly arranged in the material, which is of course true. The calculation of actual line shapes is very difficult even in the simplest cases. In the transition region above -30°c the considerable variation in line shape with temperature might well be an indication that the motion involved is complex, as indeed is to be expected, and that therefore our analysis using a single relaxation time in this region is over-simplified. Measurements of spin–lattice relaxation time will clarify this point.

3.3. *Swelled Polyisobutylene*

Experiments on the swelled polymer were made to confirm that the unusually high second moments at low temperatures were due to interactions within the molecule rather than between neighbours. They are also interesting because of the effect expected on the rubber to glass transition. Results are given in figures 1 (*b*), 2 (*b*) and 5 (*b*), for PIB swelled with benzene and having 4·88 monomers of PIB per molecule of benzene.

It is difficult to estimate the rigid lattice second moment for benzene dissolved in PIB and we use the same value as in pure benzene, namely 11·4 oersted^2 (Andrew 1950). If we then assume that the second moment of PIB is unaffected by the presence of the benzene, as would be the case if almost all the external broadening is intramolecular, one finds for 4·88 monomers PIB per C_6H_6, $\Delta H_2{}^2 = 37{\cdot}0$ oersted^2. This compares well with the experimental value of 37·6 oersted^2 (figure 2 (*b*)). Thus although the benzene has a considerable effect on the mobility of the chains in the rubbery region, which is discussed below, it hardly affects the signal from the PIB molecules at low temperatures. The peaks with small separation at -196°c (figure 1 (*b*)) are no doubt due to the benzene molecules. Since the separation (6·2 oersteds) is different from that for pure benzene (9 oersteds) it seems that little or no benzene has crystallized out on cooling the swelled polymer.

A more pronounced effect is that the principal fall in second moment occurs some 30° lower than in pure PIB and this may be ascribed to the increased freedom of movement of the PIB chains when the polymer is swelled by benzene. A similar effect has been observed in Hycar rubber swelled in benzene (Holroyd *et al.* 1951). It would again be interesting to see if a similar effect is found in the mechanical properties.

The analysis of the fall in second moment with temperature for the swelled polymer also corresponds to a thermally activated process as shown in figure 5 (*b*). We find

$$\Delta E = 0{\cdot}55 \text{ ev or } 12{\cdot}6 \text{ k cal mole}^{-1} \text{ and } \nu_0 = 4{\cdot}2 \times 10^{16} \text{ c/s (or } \Delta S = 22{\cdot}3 \text{ e.u.).}$$

The main effect of swelling is therefore a change in the frequency factor, or the entropy of activation of the motion responsible for the line narrowing.

Some preliminary experiments have been made for other concentrations of benzene and for swelling by carbon tetrachloride, which show similar effects. Nuclear magnetic resonance may be a powerful method of investigating the swelling process since, to some extent, the condition of both the solvent and the solute molecules can be studied. Even more promising would be the study of rubbers swelled with fluorocarbons since then solute and solvent motion could be readily distinguished.

ACKNOWLEDGMENTS

I am particularly grateful to Dr. A. H. Willbourn of I.C.I. Plastics Division, Welwyn Garden City, for providing the samples of polyisobutylene and for several helpful discussions; also to Dr. D. R. Holmes for providing details of his structure analysis of crystalline PIB, prior to publication.

This work was done during the tenure of an I.C.I. Fellowship and with every encouragement from Professor W. E. Curtis.

REFERENCES

ANDREW, E. R., 1950, *J. Chem. Phys.*, **18,** 607 ; 1953, *Phys. Rev.*, **91,** 425.
BLOEMBERGEN, N., PURCELL, E. M., and POUND, R. V., 1948, *Phys. Rev.*, **73,** 679.
BRILL, R., and HALLE, F., 1938, *Naturwissenschaften*, **26,** 12.
BUNN, C. W., 1942, *Proc. Roy. Soc.* A, **180,** 82 ; 1947, *J. Chem. Soc.*, p. 297.
COULSON, C. A., 1948, *Victor Henri Commemorative Volume*, p. 15 (Liège : Desoer)
COX, H. L., Jr., 1953, *Rev. Sci. Instrum.*, **24,** 307.
DICKE, R. H., 1950, *Rev. Sci. Instrum.*, **22,** 254.
EVANS, A. G., and POLANYI, M., 1943, *Nature, Lond.*, **152,** 738.
FLORY, J., 1937, *J. Amer. Chem. Soc.*, **59,** 241.
FULLER, C. S., FROSCH, C. J., and PAPE, N. R., 1940, *J. Amer. Chem. Soc.*, **62,** 1905.
GLASSTONE, S., LAIDLER, K. J., and EYRING, H., 1941, *Theory of the Rate Process*, (New York : McGraw-Hill)
GUTOWSKY, H. S., MEYER, L. H., and MCCLURE, R. E., 1953, *Rev. Sci. Instrum.*, **24,** 644.
GUTOWSKY, H. S., and PAKE, G. E., 1950, *J. Chem. Phys.*, **18,** 162.
HOLROYD, L. V., CODRINGTON, R. S., MROWCA, B. A., and GUTH, E., 1951, *J. Appl. Phys.*, **22,** 696.
HOPKINS, N. J., 1949, *Rev. Sci. Instrum.*, **20,** 401.
KAUZMANN, W., 1942, *Rev. Mod. Phys.*, **14,** 12.
LIQUORI, A. M., 1955, *Acta Cryst.*, Camb., **8,** 345.
MARVIN, R. S., 1954, *2nd Int. Congr. Rheol.* (London : Butterworth), p. 156.
PAULING, L., 1940, *Nature of the Chemical Bond*, (Cornell : University Press), p. 189
POWLES, J. G., 1956, *J. Sci. Instrum.*, in the press.
POWLES, J. G., and GUTOWSKY, H. S., 1953, *J. Chem. Phys.*, **21,** 1695.
THOMAS, P. A., and ROBINSON, D. W., 1955, *Brit. J. Appl. Phys.*, **6,** 41.
TORKINGTON, P., 1951, *J. Chem. Phys.*, **19,** 528.
TUTTLE, W. N., 1940, *Proc. Inst. Radio Engrs, N.Y.*, **28,** 23.
VAN VLECK, J. H., 1948, *Phys. Rev.*, **74,** 1168.
WALLMAN, H., MACNEE, A. B., and GADSDEN, C. P., 1948, *Proc. Inst. Radio Engrs, N.Y.*, **36,** 700.

On the Description of Rate Processes by Means of Relaxation Times

By R. O. DAVIES† and J. LAMB‡

† Physics Department, Queen Mary College, London.

‡ Electrical Engineering Department, Imperial College, London.

MS. received 19*th August* 1955 *and in amended form* 18*th October* 1955

Abstract. Attention is drawn to the importance of recognizing the various relaxation times associated with a single relaxation mechanism. The physical origin of the differences between the relaxation times is indicated and the relationships between them are illustrated by examples. A general formula is derived which relates the relaxation times for a single mechanism under different thermodynamic constraints. Connection is established between chemical rate constants and relaxation times. It is concluded that there is no intrinsic relaxation time derivable on a kinetic basis without reference to the thermodynamic conditions under which the relaxation takes place.

---

§ 1. INTRODUCTION

THE impression is often given in the literature that a relaxation mechanism possesses a relaxation time in some absolute sense. In fact, as we shall show, it is necessary to specify the independent variables with respect to which the relaxation time is defined. Neglect of this circumstance can lead to apparent inconsistencies (Manes 1953). In view of this it seems worth while to collect some essential ideas on this topic.

Although we were led to this problem through an interest in ultrasonic relaxation and the examples which we have chosen are of special interest in this field, the basic procedures are not restricted and can be applied to any single relaxation mechanism. The simplest example is, of course, the perturbation of the equilibrium state of a single chemical reaction. However, with minor adaptations, the arguments can be applied to dielectric relaxation: indeed Böttcher (1952, p. 357) has encountered a similar problem, although his case involved a linear change of thermodynamic variables whereas we shall not be restricted in this way. The basic relations between some important relaxation times were given for the first time (to our knowledge) by Davies and Jones (1953, p. 403).

In the following we shall attempt to use the term 'mechanism' to refer in a general way to a relaxation phenomenon and to confine the term 'process' to refer to a change taking place along a prescribed thermodynamic path (for example, we shall often be interested in processes in which two thermodynamic variables—such as p and T—are held fixed). We emphasize again that our work is subject to two important restrictions: (*a*) we consider only a single relaxing mechanism, (*b*) we confine our attention to variations which are sufficiently near a fixed equilibrium for the system to be treated as linear.

§ 2. THERMODYNAMIC EQUATIONS OF MOTION

Consider a simple thermodynamic system containing a single relaxation mechanism. Taking p, T as independent variables and denoting the equilibrium values of the isothermal compressibility, the specific heat and the expansivity by $\bar{\kappa}$, $\bar{C}_p$ and $\bar{\alpha}$ respectively, we have for an equilibrium change of state

$$\delta V = -V\bar{\kappa}\delta p + V\bar{\alpha}\delta T, \qquad \delta S = -V\bar{\alpha}\delta p + (\bar{C}_p/T)\delta T. \qquad \ldots\ldots(1)$$

For a very rapid change of state the thermodynamic coefficients will take their 'frozen' or 'infinite frequency' values (for which we drop the bar), the form of equations (1) remaining otherwise unaltered.

For a general small sinusoidal change in p and T it can be shown that the response is given by

$$\left.\begin{aligned} \delta V^* &= -V\left[\kappa + \frac{\Delta\kappa}{1+i\omega\tau}\right]\delta p^* + V\left[\alpha + \frac{\Delta\alpha}{1+i\omega\tau}\right]\delta T^* \\ \delta S^* &= -V\left[\alpha + \frac{\Delta\alpha}{1+i\omega\tau}\right]\delta p^* + \frac{1}{T}\left[C_p + \frac{\Delta C_p}{1+i\omega\tau}\right]\delta T^* \end{aligned}\right\} \qquad \ldots\ldots(2)$$

where the asterisks denote the usual 'complex' values, τ is a rate constant and $\Delta\kappa = \bar{\kappa} - \kappa$, etc.

The complex coefficients appearing in these expressions have a standard form, for example

$$\kappa^* = \kappa + \Delta\kappa/(1+i\omega\tau). \qquad \ldots\ldots(3)$$

The question now arises, what happens if variables other than p and T are chosen as independent? The answer can be understood by reference to a special case derived from the above. Consider the relation between p and V at constant T. Then (dropping the ω's), we have:

$$\delta V + \tau\delta\dot{V} + V\bar{\kappa}\delta p + \tau V\kappa\delta\dot{p} = 0 \quad (T = \text{constant}). \qquad \ldots\ldots(4)$$

This can be written as

$$\delta p + \frac{\kappa\tau}{\bar{\kappa}}\delta\dot{p} + \frac{1}{V\bar{\kappa}}\delta V + \frac{\kappa\tau}{\bar{\kappa}}\delta\dot{V} = 0 \quad (T = \text{constant}) \qquad \ldots\ldots(5)$$

which shows that the 'complex' bulk modulus is

$$\frac{1}{\bar{\kappa}} + \frac{\Delta(1/\kappa)}{1+i\omega(\kappa\tau/\bar{\kappa})}.$$

Generalizing this procedure, it is possible to find a set of equations similar to (2) with the notable difference that each denominator in the complex coefficients includes $\kappa\tau/\bar{\kappa}$ instead of τ.

If we now regard (3) as a standard form of coefficient then we must distinguish between relaxation times appropriate to independent variables (p, T) on the one hand, and (V, T) on the other. Therefore denoting the independent variables in the style of $\tau_{p,T} \equiv \tau$, it follows that

$$\tau_{p,T} = \frac{\bar{\kappa}}{\kappa}\tau_{V,T}. \qquad \ldots\ldots(6)$$

It can easily be shown that

$$\tau_{p,T} = \frac{\bar{\kappa}}{\kappa}\tau_{V,T} = \frac{\bar{C}_p}{C_p}\tau_{p,S} = \frac{\bar{\kappa}\bar{C}_V}{\kappa C_V}\tau_{V,S} = \frac{\bar{\alpha}}{\alpha}\tau_{p,V} = \frac{\bar{\alpha}}{\alpha}\tau_{T,S}. \qquad \ldots\ldots(7)$$

Further, it is clear from the form of equation (2) that any ordinary identity between thermodynamic coefficients (e.g. $\kappa C_V = \kappa_S C_p$, where κ_S is the adiabatic compressibility) subsists between the corresponding 'complex' or dynamic coefficients, *provided* that the appropriate relaxation times are used for each coefficient appearing in the identity.

§ 3. AN EXAMPLE

To illustrate the last remark we shall verify in detail that $\kappa^* C_V{}^* = \kappa_S{}^* C_p{}^*$. We have therefore to prove that

$$\left[\kappa + \frac{\Delta\kappa}{1+i\omega\tau_{p,T}}\right]\left[C_V + \frac{\Delta C_V}{1+i\omega\tau_{V,T}}\right] = \left[\kappa_S + \frac{\Delta\kappa_S}{1+i\omega\tau_{p,S}}\right]\left[C_p + \frac{\Delta C_p}{1+i\omega\tau_{p,T}}\right]. \qquad \ldots\ldots(8)$$

We know that

$$\kappa C_V = \kappa_S C_p \qquad \text{and} \qquad \bar{\kappa}\bar{C}_V = \bar{\kappa}_S \bar{C}_p. \qquad \ldots\ldots(9)$$

Using equation (7) $\tau_{p,S}$ and $\tau_{V,T}$ may be written in terms of $\tau_{p,T}$. With the help of (9) equation (8) may then be reduced to

$$C_V \Delta\kappa + \kappa\Delta C_V + \Delta\kappa\Delta C_V = C_p \Delta\kappa_S + \kappa_S \Delta C_p + \Delta\kappa_S \Delta C_p$$

or

$$\Delta(C_V \kappa) = \Delta(C_p \kappa_S). \qquad \ldots\ldots(10)$$

Equation (10) is, however, an identity because of (9), and this completes the verification.

The work here may be used to resolve the inconsistency noticed by Manes (1953) after his equation (27).

§ 4. THE REACTION MODEL OF A RELAXATION MECHANISM

It is always possible to describe a single relaxation mechanism by means of an internal parameter. Let this ordering parameter be denoted by Z. In the simplest (chemical) case Z may be taken as the degree of reaction so that a change in the number of molecules of species i due to the reaction is given by $dN_i = \nu_i dZ$, where ν_i is the ith stoichiometric coefficient.† In this case, also, Z can be expressed in terms of the mole numbers and the equilibrium constant for any specific example.

When a relaxation mechanism is regarded from this point of view one is tempted to argue that there must be associated with it a natural or intrinsic relaxation time τ defined by

$$dZ/dt = -(Z - \bar{Z})/\tau. \qquad \ldots\ldots(11)$$

Here $\bar{Z}$ is the equilibrium value of Z, that is, the value that would be attained if the reaction were allowed to proceed to equilibrium from its current state. It is assumed that the departure from equilibrium is sufficiently small for such a linear relation to be valid.

It is quite true that at a given instant, when the system is not in equilibrium, the value of dZ/dt is a well-defined quantity, i.e. it is in principle measurable and unique. However, the important point is that the definition of $\bar{Z}$ just given is not sufficiently precise. The value of $\bar{Z}$ to which the reaction will proceed from any point near equilibrium depends on what restraints are placed on the system. If p and T are held fixed V will (in general) change and Z will reach a certain value $\bar{Z}_{p,T}$, while if V and T are fixed it will reach $\bar{Z}_{V,T}$. *These two quantities need*

---

† By a 'single mechanism' in this case we mean that A→B symbolizes the mechanism of the reaction A→B and *not*, for example, A→C→B.

not be equal. Therefore $Z-\overline{Z}_{p,T}\neq Z-\overline{Z}_{V,T}$, and hence, since dZ/dt is fixed, the τ value must be changed to accommodate the different choices. Thus we must be specific and define $\tau_{X,Y}$ by

$$dZ/dt=-(Z-\overline{Z}_{X,Y})/\tau_{X,Y}, \qquad \ldots\ldots(12)$$

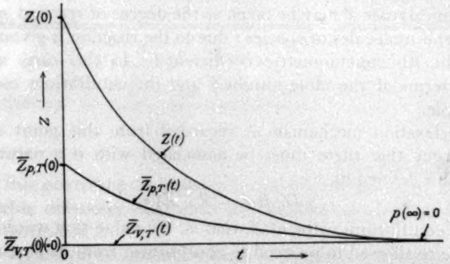
where X and Y are any pair of thermodynamic variables. The relations between the $\tau_{X,Y}$ for some possible choices of X and Y are just those of equation (7). Since (12) is true instantaneously it is the general rate equation valid even for processes in which X and Y vary with time. In this eventuality $\overline{Z}_{X,Y}$ will also vary with time although $\tau_{X,Y}$ is constant.

§ 5. THE PHYSICAL ORIGIN OF THE DIFFERENT RELAXATION TIMES

The existence of different relaxation times is caused physically by the fact that changes in the internal parameter responsible for the relaxation cause changes in the thermodynamic variables. This can be illustrated by considering the return to equilibrium of a system which is initially slightly displaced. Suppose that the volume and temperature are held fixed. Under these conditions the change in value of Z will induce a change in pressure owing to the fact that the pressure change of the reaction $(\partial p/\partial Z)_{V,T}$ is not in general zero. In fact this expression is simply proportional to the more commonly used reaction parameter $(\partial V/\partial Z)_{p,T}$ (often denoted by ΔV):

$$\left(\frac{\partial p}{\partial Z}\right)_{V,T}=\frac{1}{V\kappa}\left(\frac{\partial V}{\partial Z}\right)_{p,T}. \qquad \ldots\ldots(13)$$

To fix ideas, let the final equilibrium pressure and the final value of Z be taken as zero.



The return to equilibrium of a relaxing system at fixed volume and temperature.
(The exponential relaxation times of both Z (t) and $\overline{Z}_{p,T}(t)$ is $\tau_{V,T}$.)

On the figure we have sketched the exponential decline of Z together with that of $\overline{Z}_{p,T}$ and have indicated the value of $\overline{Z}_{V,T}(=0)$. The relaxation times of both curves $Z(t)$ and $\overline{Z}_{p,T}(t)$ are each equal to $\tau_{V,T}$: their equality arises from the linearity assumption and will appear explicitly in equation (16). The value of $\overline{Z}_{\ ,T}$ at any instant is that value of Z which would be attained if at that instant the constraint on the system were changed from constant V and T to constant p and T.

We may check the validity of the whole argument by recovering equation (6) from a description of this particular example. In the first place, by the work that led to (13) we find that

$$p=\frac{1}{V\kappa}\left(\frac{\partial V}{\partial Z}\right)_{p,T}Z. \qquad \ldots\ldots(14)$$

On the other hand if e denotes a derivative taken at equilibrium we deduce from

$$\left(\frac{\partial V}{\partial p}\right)_{e}=\left(\frac{\partial V}{\partial p}\right)_{Z,T}+\left(\frac{\partial V}{\partial Z}\right)_{p,T}\left(\frac{\partial Z}{\partial p}\right)_{e}$$

that

$$\frac{d\overline{Z}}{dp}=-\frac{V\Delta\kappa}{(\partial V/\partial Z)_{p,T}}$$

(a general relationship) which, with our special choice of origins, yields

$$\overline{Z}_{p,T}=-\frac{V\Delta\kappa}{(\partial V/\partial Z)_{p,T}}p. \qquad \ldots\ldots(15)$$

Combining equation (14) with (15) we find

$$\overline{Z}_{p,T}=-\frac{\Delta\kappa}{\kappa}Z. \qquad \ldots\ldots(16)$$

Now equation (12) can be written in the double form

$$\frac{dZ}{dt}=-\frac{(Z-\overline{Z}_{p,T})}{\tau_{p,T}}=-\frac{(Z-\overline{Z}_{V,T})}{\tau_{V,T}}\left(=-\frac{Z}{\tau_{V,T}}\text{ in our case}\right). \qquad \ldots\ldots(17)$$

Comparing equations (16) and (17) we readily recover equation (6).

§ 6. THE CONNECTION BETWEEN RATE CONSTANTS AND RELAXATION TIMES

It might perhaps be argued that a consideration of rate constants would yield a 'kinetic' relaxation time of fundamental importance. This is not the case and we shall show that what is normally regarded chemically as the relaxation time arises from a special choice of thermodynamic variables, together with a certain assumption about the ratio of the rate constants. At the same time we shall prove a general transformation formula, of which the equations in (7) are special cases.

Consider a single mechanism involving several reactants and products described by the stoichiometric coefficients ν_i. Let N_i be the mole numbers of species i and define:

$$\nu=\sum\nu_i,\quad N=\sum N_i,\quad x_i=N_i/N,\quad c_i=N_i/V,$$
$$\nu_i^+=\tfrac{1}{2}(|\nu_i|+\nu_i),\quad \nu_i^-=\tfrac{1}{2}(|\nu_i|-\nu_i),\quad \Pi_x=\Pi_i x_i^{\nu_i},\quad \Pi_c=\Pi_i c_i^{\nu_i} \qquad \ldots\ldots(18)$$
$$\Pi_x^+=\Pi_i x_i^{\nu_i+}\text{ (and similarly for }\Pi_x^-,\ \Pi_c^+,\ \Pi_c^-).$$

As previously, introduce the degree of reaction $Z(dN_i=\nu_i dZ)$ so that from the definitions (18) we have:

$$\left.\begin{aligned} d(\ln\Pi_x)&=\left(\sum\nu_i^2/N_i-\nu^2/N\right)dZ\\ d(\ln\Pi_c)&=\left(\sum\nu_i^2/N_i\right)dZ-(\nu/V)dV.\end{aligned}\right\} \qquad \ldots\ldots(19)$$

Let there be q thermodynamic variables $(\alpha_1, \alpha_2, \ldots\alpha_q)$. Then an arbitrary function of state can be written as $F(\alpha_1,\ldots\alpha_q, Z)$. In equilibrium Z takes up a value $\overline{Z}(\alpha_1,\ldots\alpha_q)$ and F becomes $\overline{F}(\alpha_1,\ldots\alpha_q)\equiv F[\alpha_1,\ldots\alpha_q, \overline{Z}(\alpha_1,\ldots\alpha_q)]$. It is easy to show that

$$(\partial F/\partial\alpha_l)-(\partial F/\partial\alpha_l)\equiv\Delta(\partial F/\partial\alpha_l)=(\partial F/\partial Z)_{\alpha_l}(\partial\overline{Z}/\partial\alpha_l). \qquad \ldots\ldots(20)$$

We consider three possible assumptions concerning the form of the rate equation:

$$\left.\begin{aligned} &\text{(A)}\quad \frac{1}{V}\frac{dZ}{dt} = k_c^{+}\Pi_c^{-} - k_c^{-}\Pi_c^{+},\\ &\text{(B)}\quad \frac{1}{V}\frac{dZ}{dt} = k_x^{+}\Pi_x^{-} - k_x^{-}\Pi_x^{+},\\ &\text{(C)}\quad \frac{dZ}{dt} = LA \end{aligned}\right\} \qquad \ldots\ldots(21)$$

where the rate constants k_c^{+} etc. and L are functions of state, A is the affinity of the reaction $(= -\Sigma\nu_i\mu_i)$ and in equilibrium $\bar{A}=0$, $\bar{k}_c^{+}\bar{\Pi}_c^{-} = \bar{k}_c^{-}\bar{\Pi}_c^{+}$, and $\bar{k}_x^{+}\bar{\Pi}_x^{-} = \bar{k}_x^{-}\bar{\Pi}_x^{+}$.

Introduce, as functions of state, $K_c = k_c^{+}/k_c^{-}$ and $K_x = k_x^{+}/k_x^{-}$; evidently $\bar{K}_x = \bar{\Pi}_x$ and $\bar{K}_c = \bar{\Pi}_c$. Let the equations of (21) be expanded to the first order about a fixed equilibrium state. Then:

$$\left.\begin{aligned} &\text{(A)}\quad \frac{1}{V}\frac{d\delta Z}{dt} = \bar{k}_c^{-}\bar{\Pi}_c^{+}[\delta \ln (K_c/\Pi_c)],\\ &\text{(B)}\quad \frac{1}{V}\frac{d\delta Z}{dt} = \bar{k}_x^{-}\bar{\Pi}_x^{+}[\delta \ln (K_x/\Pi_x)]\\ &\text{(C)}\quad \frac{d\delta Z}{dt} = L\delta A. \end{aligned}\right\} \qquad \ldots\ldots(22)$$

In order to find the relaxation times for processes in which the α_l are fixed, imagine the differentials on the right of (22) expanded in terms of the $\delta\alpha_l$ and δZ. Taking the terms in δZ and $d\delta Z/dt$ as the only variable terms we see that:

$$\left.\begin{aligned} &\text{(A)}\quad \tau_{\alpha_l}^{-1} = -V\bar{k}_c^{-}\bar{\Pi}_c^{+}[\partial \ln (K_c/\Pi_c)/\partial Z]_{\alpha_l}\\ &\text{(B)}\quad \tau_{\alpha_l}^{-1} = -V\bar{k}_x^{-}\bar{\Pi}_x^{+}[\partial \ln (K_x/\Pi_x)/\partial Z]_{\alpha_l}\\ &\text{(C)}\quad \tau_{\alpha_l}^{-1} = -L(\partial A/\partial Z)_{\alpha_l}. \end{aligned}\right\} \qquad \ldots\ldots(23)$$

Consider a change of thermodynamic variables to $\beta_1, \ldots \beta_q$, where $\beta_l = f_l(\alpha_1, \ldots \alpha_q, Z)$. For any function of state F we may show (with the help of equation (20)) that:

$$\left(\frac{\partial F}{\partial Z}\right)_{\alpha_l} = \left(\frac{\partial F}{\partial Z}\right)_{\beta_l} + \sum_m \frac{\partial \bar{F}}{\partial \beta_m}\frac{\partial \beta_m}{\partial Z} - \left(\frac{\partial F}{\partial Z}\right)_{\beta_l}\sum_{m,n}\frac{\partial \bar{\alpha}_m}{\partial \beta_n}\Delta\left(\frac{\partial \beta_n}{\partial \alpha_m}\right). \qquad \ldots\ldots(24)$$

Now, in equation (24), place F equal to $\ln (K_c/\Pi_c)$, $\ln (K_x/\Pi_x)$, and A in turn. In each case $\bar{F} \equiv 0$ so that, for these three functions F, we have

$$\left(\frac{\partial F}{\partial Z}\right)_{\alpha_l} = \left(\frac{\partial F}{\partial Z}\right)_{\beta_l}\left[1 - \sum_{m,n}\left(\frac{\partial \bar{\alpha}_m}{\partial \beta_n}\right)\Delta\left(\frac{\partial \beta_n}{\partial \alpha_m}\right)\right]. \qquad \ldots\ldots(25)$$

Finally from equation (23) we see that for each case (A), (B) and (C):

$$\tau_{\beta_l} = \tau_{\alpha_l}\left[1 - \sum_{m,n}\left(\frac{\partial \bar{\alpha}_m}{\partial \beta_n}\right)\Delta\left(\frac{\partial \beta_n}{\partial \alpha_m}\right)\right]. \qquad \ldots\ldots(26)$$

This is a general formula which relates the relaxation times for a single mechanism under different thermodynamic constraints.

As an example of the application of equation (26) we shall again obtain equation (6). Take $(\beta_1, \beta_2)=(p, T)$ and $(\alpha_1, \alpha_2)=(V, T)$. Then

$$\tau_{p,T}=\tau_{V,T}\left[1-\left(\frac{\partial\bar{V}}{\partial p}\right)_T \Delta\left(\frac{\partial p}{\partial\bar{V}}\right)_T-\left(\frac{\partial\bar{T}}{\partial p}\right)_T \Delta\left(\frac{\partial p}{\partial\bar{T}}\right)_V-\left(\frac{\partial\bar{V}}{\partial\bar{T}}\right)_p \Delta\left(\frac{\partial T}{\partial\bar{V}}\right)_T \right.$$
$$\left.-\left(\frac{\partial\bar{T}}{\partial\bar{T}}\right)_p \Delta\left(\frac{\partial T}{\partial\bar{T}}\right)_V\right].$$

The last three terms are zero and we readily recover equation(6).

Having obtained the desired transformation formula, it is worth noting that with our definitions of the *k*'s and the *K*'s, cases (A) and (B) above are in fact equivalent. This follows from the physical equality between $k_c{}^+\Pi_c{}^-$ and $k_x{}^+\Pi_x{}^-$, for example, each representing the number of forward reactions per unit time and volume. Then, since the Π_c's and Π_x's are related through the definitions of (18), the k_c's and k_x's may also be related.

We have shown above that the relation (26) between the relaxation times is the same for each of the three (or really two) special assumptions of equations (21). As a matter of fact it can be further shown that the whole phenomenological theory of a single relaxation mechanism is independent of which rate assumption of (21) is chosen as the starting point.

In conclusion, we wish to note that, by virtue of (19), we may write the first expression for the relaxations time in (23) as:

$$\tau_{\alpha_l}{}^{-1}=V\bar{k}_c{}^-\bar{\Pi}_c{}^+\left[\sum\nu_i{}^2/N_i-(\partial\ln K_c/\partial Z)_{\alpha_l}-(\nu/V)(\partial V/\partial Z)_{\alpha_l}\right]. \quad \ldots\ldots(27)$$

Consider the following simple assumptions which could be made:

(i) $(\partial\ln K_c/\partial Z)_{V,T}=0,$

(ii) $(\partial\ln K_c/\partial Z)_{p,T}=0.$

Applying these conditions in turn to equation (27) we find

(i) $\tau_{V,T}{}^{-1}=V\bar{k}_c{}^-\bar{\Pi}_c{}^+\left[\sum\nu_i{}^2/N_i\right],$

(ii) $\tau_{p,T}{}^{-1}=V\bar{k}_c{}^-\bar{\Pi}_c{}^+\left[\sum\frac{\nu_i{}^2}{N_i}-\frac{\nu}{V}\left(\frac{\partial V}{\partial Z}\right)_{p,T}\right].$

With the further specialization to the reaction $\nu_1=1$, $\nu_2=-1$, we find that

(i) $\tau_{V,T}{}^{-1}=\frac{V}{N}(\bar{k}_x{}^+ + \bar{k}_x{}^-)=(\bar{k}_c{}^+ + \bar{k}_c{}^-),$

(ii) $\tau_{p,T}{}^{-1}=(\bar{k}_c{}^+ + \bar{k}_c{}^-).$

These assumptions are in general mutually incompatible, and no particular significance can be attached to the above deduction since both the nature of the reaction and the dependence of K_c on Z have to be assumed in order to derive a relaxation time equal to $1/(\bar{k}_c{}^+ + \bar{k}_c{}^-)$.

§ 7. CONCLUSIONS

We conclude that there is a multiplicity of relaxation times associated with a single parameter relaxation mechanism. These times are distinguished by the independent variables to which they refer; they are related by the general formulae (26), specific examples of which are given in equation (7), and must be chosen appropriately in the formation of time-dependent thermodynamic coefficients.

Further, there is no intrinsic relaxation time (e.g. one derived kinetically) which can be singled out in any way other than by the choice of variables just mentioned. In order to identify a kinetically derived relaxation time it is therefore necessary to scrutinize the work to decide which two thermodynamic variables have been used in the derivation.

No particular significance can be attached to a relaxation time equal to $(\bar{k}_c^{+}+\bar{k}_c^{-})^{-1}$: this value is only obtained under fixed conditions when specific assumptions are made concerning the nature of the reaction and the dependence of K_c on Z.

REFERENCES

Bøttcher, C. J. F., 1952, *Theory of Electric Polarisation* (Amsterdam : Elsevier.)
Davies, R. O., and **Jones**, G. O., 1953, *Advanc. Phys.*, *Phil. Mag.* Suppl., **2,** 370.
Manes, M., 1953, *J. Chem. Phys.*, **21,** 1791.

Effective Tilts of the Ionosphere at Places about 1000 km apart

By H. A. WHALE

Seagrove Radio Research Station, Auckland University College, New Zealand

Communicated by P. W. Burbidge; MS. received 8th *March* 1955 *and in amended form* 25th *August* 1955

Abstract. A comparison of the bearing and elevation angles of radio waves from two stations situated about 2000 km to the North and to the West of the receiving site yields results which are sometimes consistent with the existence during the daytime of effective ionospheric tilts which are similar at places more than 1000 km apart. These tilts are of the order of half a degree and are in opposite directions during the morning and the afternoon. They may correspond to actual tilts of a uniform ionosphere or be due to horizontal gradients of electron density. The factor relating the tilt angle to the observed bearing deviations was measured experimentally and the effective tilt angle found to be approximately twice the bearing deviation.

§ 1. INTRODUCTION

DURING the period July 1953 to June 1954 a series of measurements was made of the angles of arrival of radio waves from two stations at about the same distance but with different bearings from the receiving site. The positions of these two transmitting stations and the receiving site are shown in figure 1 *a*, which is a portion of a great-circle map (equidistant azimuthal projection) centred on Seagrove, Auckland. The two transmitting stations are ZQD, Nandi, Fiji, operating on 9315 kc/s and distant 2000 km on a great circle bearing of about 90° East of North, and VLQ9, Brisbane, Queensland, operating on 9660 kc/s and distant 2250 km on a great circle bearing of about 291° East of North.

The measurements bearing and elevation were made with the rotating radio interferometer which has been described previously (Whale 1954). In all cases these measurements were made on the horizontally polarized component of the incoming wave, the signals being received on horizontal multiple-loop aerials. As the ground-reflected wave is almost out of phase with the direct wave at low elevation angles the radio signal from the aerial is relatively weak for waves arriving at low angles. When there is only one main direction of incoming signal, there is no uncertainty as to which direction is being measured but, when there are signals arriving simultaneously from a variety of different elevations the measured direction is nearest that of the particular incoming wave which produces the greatest output from the aerial. This ground effect, which is always present when measurements are made on the horizontally polarized component of a radio wave, is almost eliminated when the measurements are made on the vertically polarized component. However, since it is usual to use horizontally polarized aerials for short-wave radio communication the measurements described here are confined to this component. It has been found that, when vertical aerials are used, the measured elevation angles are often smaller than those obtained with horizontal aerials.

The aim of this investigation was the measurement of the effects of regular tilts and their experimental magnitude compared with other disturbances as a step towards an explanation of the off-great-circle propagation which may occur over long distances.

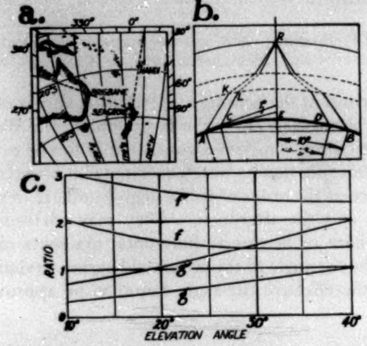


Figure 1. *a*. Portion of equidistant azimuthal map centred on Seagrove with bearings in degrees East of true North.

b. Diagrammatic representation of the two transmission paths Nandi–Seagrove (CD) and Brisbane–Seagrove (AB).

c. Ratio of the observed *change* in angle to the effective tilt angle of the ionosphere as a function of the elevation angle of the received wave for bearing angle (one hop, g; two hops, g') and elevation angle (one hop, f; two hops, f').

§ 2. Diurnal Variations of the Angle of Arrival

The measured values of the bearing angles and the elevation angles have been averaged over each hourly period centred on the half-hour. Some representative measured values for March 1954 are plotted in figure 2. Each point on the graphs represents the reading averaged over one hour and some typical successive readings have been connected by a thin line. The accuracy of the measurements is closer than ± 2° for elevation angles above 10° but only to about ± 5° below that. The bearing angles are measured to the nearest degree. The bearings shown are in degrees East of true North while the times are the local times applicable to the mid-point of the path.

All the measured values have then been averaged over each month for each hourly period. The results have been plotted in the form shown in figures 3 *a* and 3 *b* where deviation of the monthly average bearing angle from the overall average (great circle) bearing for each hourly period has been plotted as a function of the monthly average elevation angle for signals received from stations ZQD and VLQ9 for March 1954. The heavy lines are smoothed curves through the experimental points and the times are again the local times at the mid-points of the propagation paths. This method of presenting the experimental results has been adopted since it shows the range of angles which a directional receiving aerial must cover. These two graphs are reproduced on a reduced scale in the series shown as figure 4 where all the results obtained over the period July 1953 to June 1954 are plotted. The star in each graph indicates midday while the arrow indicates the direction on the locus of increasing time.

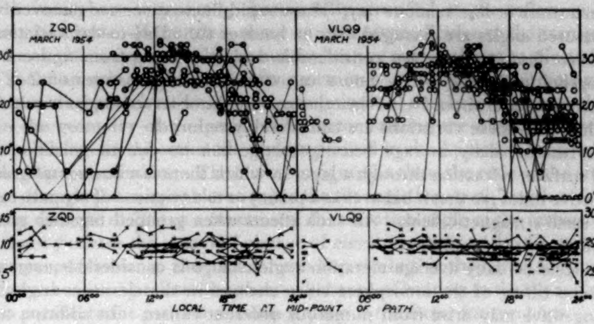


Figure 2. Representative observed elevation and bearing angles averaged over each hourly period for March 1954.

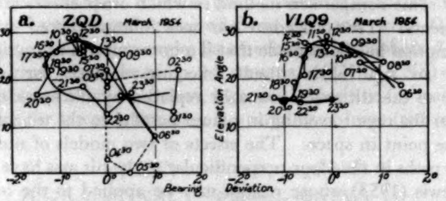


Figure 3. Observed elevation angles and bearing angles averaged over each month for each hourly period plotted as a diurnal variation in space of the direction of arrival.

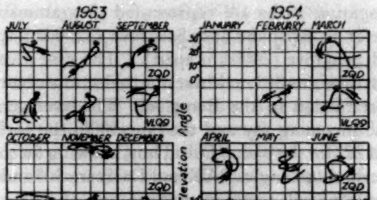


Figure 4. Averages over each month of the diurnal variations of the direction of arrival of signals from ZQD and VLQ9.

§ 3. COMPARISON OF RESULTS FOR THE TWO STATIONS

It is well known that the ionosphere does not usually behave as a smooth reflecting surface but has a certain degree of roughness so that the incoming wave is made up of an aggregate of waves which cluster around the mean direction. The measured angles (both bearing and elevation) thus fluctuate about the mean angle in a more or less random way. These rapid fluctuations are removed when the measured angles are averaged over an hour or so. Day-to-day variations in ionosphere characteristics are considerable and it appears from figure 2 that inhomogeneities lasting a few hours can cause bearing changes of 2–3° and elevation changes of about 5°. When the mean of the observed angles is calculated for each month these variations are considerably reduced.

The residual daily average bearing change can be due to reflection from a tilted surface, refraction through a layer in which there is a horizontal gradient of refractive index (so that it behaves as a prism) or to deviations of ray path caused by the earth's magnetic field. All such effects when grouped together give the effective tilt of the ionosphere.

The residual daily average elevation angle change is considerable even when there is no tilting of the ionosphere since changes in the elevation angle of an incoming wave may arise from numerous different causes. In addition to the tilt effect, the height of the reflecting layer may change, the wave may be refracted by different amounts in both the underlying layers and in the reflecting layer and propagation may be by different numbers of hops. The effects of tilts can only be isolated by some comparison method in which other effects are removed from the measurements. This procedure has been followed here.

The assumption has been made that if a non-uniform refracting layer may be replaced by an equivalent tilted reflecting surface for rays in planes perpendicular to the tilt axis, it may be replaced by the same equivalent tilted reflecting layer for rays travelling in planes parallel to the tilt axis and reflected from the same point in space. The effects of two models of non-uniform ionosphere on ray paths in the plane perpendicular to the tilt axis have been calculated by Waldo Lewis (1953) whose results may be applied to the longitudinal case using the equivalent vertical-incidence frequency and the above assumption shown to be obeyed. Calculations by Keam (1954) using a different model for the ionosphere also show that the assumption is valid for all practical purposes.

The two propagation paths are represented diagrammatically in figure 1*b* where A and B refer to Brisbane and Seagrove (2250 km apart) and C and D refer to Nandi and Seagrove (2000 km apart). The case illustrated corresponds to one possible daytime condition where the reflection occurs in the F1 region with appreciable refraction in the E region. The average *measured* elevation angles between the incoming ray and the tangent to the earth's surface are 20° for the path AB and 22° for the path CD. Thus, allowing for the earth's curvature, the average measured chord angles BAK and DCL are 30° and 31°. The average angles of incidence on the reflecting region of the ionosphere are thus 60° (9660 kc/s) and 59° (9315 kc/s) for single-hop transmission, and, by a similar reasoning, 65° (9660 kc/s) and 63·5° (9315 kc/s) for double-hop transmission. Thus the equivalent vertical-incidence frequencies for these four cases are given by:

$$\begin{aligned} \text{single hop:}\quad & 60^\circ(9660\,\text{kc/s}) \equiv 0^\circ(4830\,\text{kc/s}),\\ & 59^\circ(9315\,\text{kc/s}) \equiv 0^\circ(4797\,\text{kc/s}),\\ \text{double hop:}\quad & 65^\circ(9660\,\text{kc/s}) \equiv 0^\circ(4082\,\text{kc/s}),\\ & 63{\cdot}5^\circ(9315\,\text{kc/s}) \equiv 0^\circ(4156\,\text{kc/s}). \end{aligned}$$

For each of the single-hop and the double-hop cases, reflection of the rays will occur at very nearly the same ionization density for both stations.

The ionosphere may not be horizontally stratified at the reflection points but may be tilted. If the tilt is about an axis perpendicular to the transmission path the elevation angle will change but the bearing angle will not be affected, while a tilt about an axis parallel to the transmission path will affect the bearing very much more than the elevation angle, the magnitude of the observed change depending on the geometry of the path. The factor relating the observed bearing or elevation change to the actual tilt angle is derived in the Appendix for single- and double-hop transmission. The curves in figure 1 *c* show the ratio of the observed change of angle (bearing or elevation) to the tilt angle for a path length of 2200 km (angular distance of 20°) as a function of the elevation angle of the received wave. The factor f relates the observed change of elevation angle to the tilt angle about an axis perpendicular to the path for one hop while f' is the factor for two hops. Similarly the factor g relates the observed change of bearing to the tilt angle about an axis parallel to the path for one hop while g' is the factor for two hops.

The change in angle for two-hop propagation is nearly twice the change for one-hop propagation for a given vertical angle of arrival. Let us assume that the effects of a tilt on bearing angles and on elevation angles may be considered separately. If the subscript 1 refers to the signals received from Nandi and the subscript 2 refers to the signals received from Brisbane, write $D_{1,2}$ = observed elevation angles, $\Delta_{1,2}$ = elevation angles for a horizontal ionosphere, $\theta_{1,2}$ = observed bearings, $\Theta_{1,2}$ = great circle bearings of stations, $\xi(\mathrm{EW})$, $\xi(\mathrm{NS})$ = components of the tilt angle about EW and NS axes, regarded as positive for increasing height towards the North or towards the East, n = number of hops (either one or two). Then

$$\left.\begin{array}{ll} D_1=\Delta_1+nf\xi(\mathrm{EW}) & \theta_1=\Theta_1-ng\xi(\mathrm{NS}) \\ D_2=\Delta_2-nf\xi(\mathrm{NS}) & \theta_2=\Theta_2-ng\xi(\mathrm{EW}) \end{array}\right\} \quad \ldots\ldots(1)$$

From these equations,

$$\left.\begin{array}{r} D_1-D_2=\Delta_1-\Delta_2+nf\{\xi(\mathrm{EW})+\xi(\mathrm{NS})\} \\ \theta_1+\theta_2=\Theta_1+\Theta_2-ng\{\xi(\mathrm{EW})+\xi(\mathrm{NS})\}. \end{array}\right\} \quad \ldots\ldots(2)$$

We are concerned with variations of the angles about their mean values. The measured average value of the quantity $\Delta_1-\Delta_2=2°$ while the measured average value of $\Theta_1+\Theta_2=300°$. These average values will be considered as constants so that a plot of D_1-D_2 as ordinate with $\theta_1+\theta_2$ as abscissa for pairs of measurements taken at the same time will give a curve of slope $-f/g$. Both f and g are functions of the elevation angle of the received wave and since the time of reception and the elevation angle are not linearly related the slope obtained from this plot will be the average value of $-f/g$. From figure 1 *c*, this average value (for the usual range of elevation angles of 15° to 35°) is about 0·4. Plots of this nature have been made in figure 5 *a* for the months for which results are available for both stations. The horizontal scale is 1° per division while the vertical scale is 5° per division in each case. The dashed portions of the curves refer to times between sunrise and midday, the full line portions refer to times between midday and 1800 hours local time at the midpoint of the path, while the dotted portions of the curves are for night-time results. The graphs have been drawn using points obtained for the same local time at the mid-point of the path.

The points from which the smoothed curve has been drawn are shown for the first graph only, i.e. for July 1953. A straight line of slope −0·5 has been drawn in each graph. Only the curve for May 1954 fits this straight line well. However, those for July 1953 and March, April and May 1954 are sufficiently similar to warrant further investigation. Accordingly, the averaged results for these four months are also plotted in figure 5 *b*. In this case a much better fit to the straight line is obtained.

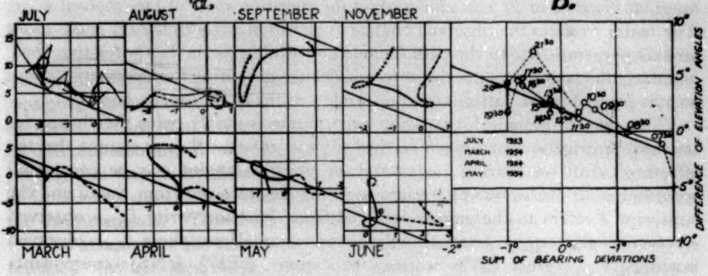


Figure 5 *a*. Average differences in observed elevation angles $D_1 - D_2$ plotted vertically as a function of the average sum of the observed bearing angles $\theta_1 + \theta_2$ horizontally for each of the months shown.

b. The same but averaged over the four months July 1953 and March, April and May 1954.

The results for July 1953 and March, April and May 1954 thus indicate the presence of widespread ionospheric tilts and yield the experimental value of 0·5 for the factor $-f/g$ relating the bearing change to the elevation change for a given ionospheric tilt. This agrees with the calculations which yield a value of 0·4 when averaged over the normal range of observed elevation angles.

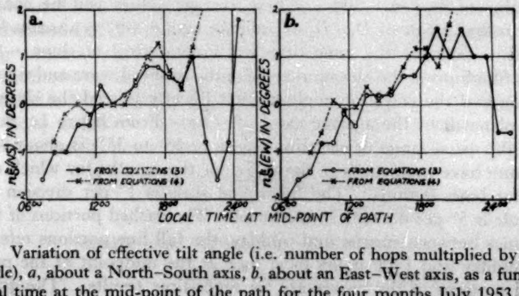


Figure 6. Variation of effective tilt angle (i.e. number of hops multiplied by average tilt angle), *a*, about a North–South axis, *b*, about an East–West axis, as a function of the local time at the mid-point of the path for the four months July 1953 and March, April and May 1954. A tilt is regarded as positive if the height increases towards the East or the North.

§4. MAGNITUDE OF TILTS

We could have taken the *observed* bearing deviations and, multiplying them by the *calculated* factors g and g' deduced the effective tilt angles about axes parallel to the transmission paths. Thus, as in equations (1)

$$\left.\begin{aligned} n\xi(\mathrm{NS}) &= -(\theta_1-\Theta_1)/g = -\phi_1/g \\ n\xi(\mathrm{EW}) &= -(\theta_2-\Theta_2)/g = -\phi_2/g \end{aligned}\right\} \quad \ldots\ldots(3)$$

where $\phi_1=\theta_1-\Theta_1$ and $\phi_2=\theta_2-\Theta_2$, the deviation of the bearing from the average bearing.

An equivalent method, which requires the elevation angles as well as the bearing angles and hence may be expected to give a better average value of the tilt angle is as follows: Write $D_1-D_2=D$, the difference of observed elevation angles, $\Delta_1-\Delta_2=\Delta=2^\circ$ on the average. Then from equation (2),

$$D-\Delta=nf\{\xi(\mathrm{EW})+\xi(\mathrm{NS})\}$$

and, from equation (3), eliminating g,

$$\phi_1/\phi_2=\xi(\mathrm{NS})/\xi(\mathrm{EW}).$$

Thus,

$$\left.\begin{aligned} n\xi(\mathrm{NS}) &= (D-\Delta)/\{f(1+\phi_2/\phi_1)\} \\ n\xi(\mathrm{EW}) &= (D-\Delta)/\{f(1+\phi_1/\phi_2)\}. \end{aligned}\right\} \quad \ldots\ldots(4)$$

The effective tilt angles, $n\xi(\mathrm{NS})$ and $n\xi(\mathrm{EW})$, calculated by both equations (3) and (4) are plotted in figure 6, for the months July 1953 and March, April and May 1954. Here, the approximate values, $f=1$ and $g=0{\cdot}5$ have been used. Good agreement between the two methods is obtained.

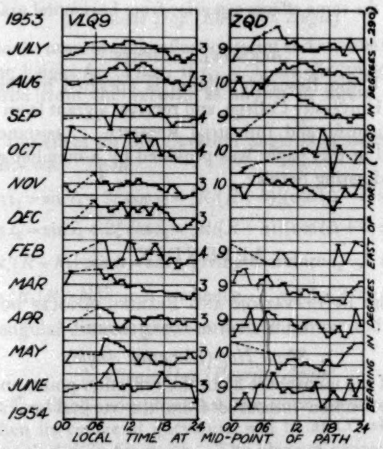


Figure 7. Observed average diurnal variation of bearing angles for the months shown.

All the average *bearing* measurements for both stations are shown in figure 7 where the bearing scale is one degree per division increasing upward. These curves are somewhat similar to those observed by Budde (1950). Although it has not been shown that the effective tilt angles derived from these bearing curves occur simultaneously over both paths in all cases it is probable that an approximation to the actual effective tilt angles about the NS and EW axes may be obtained from them by multiplying the values of bearing deviation by a factor of about two since it is does not seem that bearing deviations can arise in any other way than by reflection from an ionosphere which behaves as though it is tilted.

Miya (1954) has shown that the difference between the elevation angles of a transmitted wave (from San Francisco) and the received wave (at Fukuoka, a path length of about 8500 km) has a diurnal variation which would indicate effective tilts of the same order of magnitude as we have deduced from the shorter path measurements.

§ 5. Conclusion

From measurements of the bearing angles and the angles of elevation of radio waves from two stations, average ionospheric tilts of the order of about a degree have been shown to be similar over distances of the order of 1000 km during the months July 1953 and March, April and May 1954. During these months the effective ionospheric tilts during the daytime have been shown to be obtainable from the bearing deviations by multiplying these by a factor of about two.

The average bearing deviations for August, September and November 1953 and June 1954 are also plotted. The approximate values of the effective ionospheric tilts can be deduced from these bearings but the related measurements on the elevation angles do not indicate conclusively that the tilts were widespread during these months.

Although the results obtained in this work indicate the presence of effective tilts of the ionosphere, these are not necessarily bodily tilts of the ionosphere as a whole because the same effects can arise from horizontal gradients of electron density.

Acknowledgments

The Seagrove Radio Research Station is operated by the Department of Physics, Auckland University College, and receives a grant from the New Zealand Department of Scientific and Industrial Research. Assistance in making the measurements used in this paper was provided by a member of the staff of the New Zealand Broadcasting Service.

APPENDIX

Measured Changes in Elevation and Bearing Angles due to Effective Tilts of the Ionosphere

1. *Single Hop* (*Elevation Angle*)

If ξ is the tilt of the ionosphere at L, the reflection point, and i_1, i_2 the angles of incidence and reflection measured from the vertical as in figure 8*a*, then $i_2 - i_1 = 2\xi$.

Let
$$i_1 = i_0 - \xi, \qquad i_2 = i_0 + \xi. \qquad \text{......(A 1)}$$

Assume that the effect of a tilt on the elevation angle is symmetrical so that the angle LAK = angle KBL = δ. From the triangles OAL, OBL,

$$\left.\begin{aligned} \sin i_1 &= \{R/(R+h)\}\cos(\Delta+\delta) \\ \sin i_2 &= \{R/(R+h)\}\cos(\Delta-\delta). \end{aligned}\right\} \qquad \text{......(A 2)}$$

Both ξ and δ are small, so that from (A 1) and (A 2)

$$\sin i_1 = \sin i_0 - \xi \cos i_0 = \{R/(R+h)\}(\cos\Delta - \delta \sin\Delta)$$
$$\sin i_2 = \sin i_0 + \xi \cos i_0 = \{R/(R+h)\}(\cos\Delta + \delta \sin\Delta).$$

Hence $\sin i_0 = \{R/(R+h)\}\cos\Delta = \sin i$, and $i_0 = i$, the angle of incidence with no tilt; also $\xi \cos i = \{R/(R+h)\}\delta \sin\Delta$ whence $\delta = \xi \cot i \cot\Delta$ and $f = \delta/\xi = \tan(\Delta+\lambda)\cot\Delta$, where 2λ is the angular distance between the transmitter and the receiver.

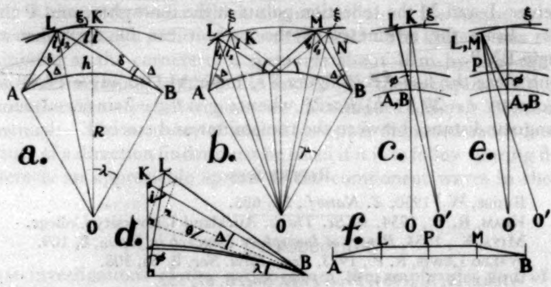


Figure 8. *a*, Change in elevation angle with tilt for one hop ; *b*, change in elevation angle with tilt for two hops ; *c* and *d*, change in bearing angle with tilt for one hop ; *e* and *f*, change in bearing angle with tilt for two hops.

2. *Double Hop* (*Elevation Angle*)

Let the tilt ξ be the same at each reflection point and assume that both reflections occur at the same height h. Then, if i_1, i_2, i_3 are the angles of incidence and reflection measured from the vertical as shown in figure 8*b*, $i_1 = i_2 - 2\xi$ and $i_3 = i_2 + 2\xi$.

Assume that the effect on the elevation angle is symmetrical so that angle LAK = angle MBN = δ. Then

$$\sin i_1 = \sin i_2 - 2\xi \cos i_2 = \{R/(R+h)\}(\cos(\Delta-\delta)\sin\Delta)$$
$$\sin i_3 = \sin i_2 + 2\xi \cos i_2 = \{R/(R+h)\}(\cos(\Delta+\delta)\sin\Delta).$$

Hence $\sin i_2 = \{R/(R+h)\}\cos\Delta = \sin i$, giving therefore $i_2 = i$, and

$$2\xi \cos i = \{R/(R+h)\}\delta \sin\Delta \quad \text{whence} \quad f' = \delta/\xi = 2\tan(\Delta+\mu)\cot\Delta$$

where 4μ is the angular distance between the transmitter and the receiver.

3. *Single Hop* (*Bearing Angle*)

Assume that the tilted ionosphere has the same curvature as the horizontal ionosphere. Then its centre of curvature will be at O′ where angle OKO′ is equal to the tilt angle ξ as in figure 8*c*. The plane of propagation will then pass through A, B and O′ and the angle ϕ between this plane and a vertical plane at the mid-point of the path is given by

$$\phi = \xi(R+h)/R \simeq \xi.$$

In figure 8*d*, angle LBK = $\phi \cos i$.

The measured bearing deviation θ is the projection of angle LBK on the earth's surface at B

i.e. $$\theta = (\text{angle LBK}) \sec\Delta = \phi \cos i \sec\Delta \simeq \xi \cos i \sec\Delta$$

whence $g = \theta/\xi = \sin(\Delta+\lambda)\sec\Delta$ where 2λ is the angular distance between the transmitter and the receiver.

4. *Double Hop* (*Bearing Angle*)

A plan view of the path will be as shown in figure 8*f* where A is the transmitter, B the receiver, L and M the reflection points at the ionosphere and P the ground reflection. From the symmetry of the path it can be seen that the angle ALP = angle BMP.

Looking along the line AB in figure 8*e*, angle ALO′ = angle PLO′ and, from simple geometry $\phi = 2\xi(R+h)/R \simeq 2\xi$ whence $g' = \theta/\xi = 2\sin(\Delta+\lambda)\sec\Delta$ where 2λ is the angular distance between the transmitter and receiver.

REFERENCES

BUDDE, W., 1950, *Z. Naturf.*, **5a**, 683.

KEAM, R. F., 1954, *M.Sc. Thesis*, Auckland University College.

MIYA, K., 1954, *Report of Ionosphere Research in Japan*, **8**, 109.

WALDO LEWIS, R. P., 1953, *Proc. Phys. Soc.* B, **66**, 308.

WHALE, H. A., 1954, *Proc. Phys. Soc.* B, **67**, 553.

An Automatic Direction Finder for recording Rapid Fluctuations of the Bearing of Short Radio Waves

BY H. A. WHALE AND W. J. ROSS†
Seagrove Radio Research Station, Auckland University College, Auckland, New Zealand

Communicated by P. W. Burbidge; MS. received 13*th May* 1955 *and in amended form* 21*st October* 1955

Abstract. A direction finder which aligns itself along the horizontal projection of the wave normal of a radio wave is described. The system uses two pairs of aerials situated at the corners of a rectangle about 10 m by 5 m, the signals in each pair of aerials being combined in such a way that a required 90° phase shift is maintained accurately at all frequencies up to about 17 Mc/s. Sense determination is manual. The accuracy of such a system is discussed and it is shown that the aperture of a direction finder must be small if it is to follow bearing fluctuations when there is an appreciable spread of the component waves in the incoming signal.

§ 1. INTRODUCTION

FROM investigations of the variations of the amplitudes and of the phase differences between the signals received on pairs of aerials it is known (Bramley 1951) that a radio wave reflected from the ionosphere usually consists of waves from a large number of different directions which may be grouped around a few main directions. The main directions generally correspond to the paths which could be calculated from a knowledge of the average ionospheric characteristics while the spreading of the waves about these directions seems to arise from scattering processes due to the existence of inhomogeneities in the ionosphere.

Rapid fluctuations of the direction of the normal to the resultant wave front, i.e. of both the instantaneous bearing and the elevation angle, are associated with the amplitude fading of the incoming radio wave. The recording direction finder described below has been built in order that the rapid fluctuations in bearing may be investigated experimentally.

§ 2. OPERATION

The instantaneous direction of the wave normal in the horizontal plane will depend on the relative amplitudes, directions and phases of the various incoming waves. It will also depend on the exact location of the direction finder although it is to be expected that similar direction finders at slightly different places observing on the same transmission will give readings which, while not identical, have similar statistical properties when taken over a period of time.

The direction of the wave normal may be found in two basically different ways. In the first of these, the phase differences between the signals received in pairs of fixed aerials are measured and the direction of the wave normal calculated from these phase differences. In methods used by Bramley and Ross (1951) and Miya and Ishikawa (1955) two pairs of spaced aerials were used, the horizontal spacing being equal in both the north–south and east–west directions.

---

† Now at the Ionosphere Research Laboratory, Pennsylvania State University.

In the second method, either the aerial system itself or a pickup coil is rotated to a null position. Most commercial direction finders, the rotating radio interferometer (Whale 1954) and the automatic direction finder described here belong to this class.

There is an inherent lag in a mechanical following system such as the one described here so that the ultimate response time cannot approach that attainable in systems using cathode-ray tube display systems. However, where a very large number of records is required, the advantages of direct recording are considerable and it also seems that this particular following system can be designed to follow the most rapid fluctuations of wave normal direction which occur in practice.

The horizontal tangent to a wave front is the line through two closely spaced identical aerials which yield signals with zero phase difference. It is difficult to determine automatically the orientation of an aerial pair to produce zero phase difference but a simple method of determining whether the phase difference between two signals is less than or greater than 90° has been described previously (Whale 1954). This method may now be used for determining when there is zero phase difference between the signals from the two aerials provided an initial fixed 90° phase change is given to one of them. For simplicity of operation this 90° phase change must be maintained accurately over the entire frequency range of the direction finder.

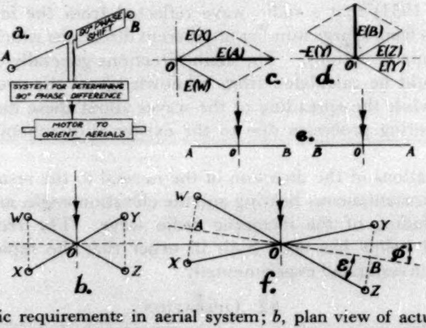


Figure 1. *a*, Basic requirements in aerial system; *b*, plan view of actual aerial array; *c*, addition of signals from aerials W and X; *d*, subtraction of signals from aerials Y and Z; *e*, the two stable positions of the aerial; *f*, aerial array off the null position.

The basic system required is then as drawn in figure 1 *a*. The fixed 90° phase shift in the signal from one aerial can be simply obtained by arranging aerials as in the plan view in figure 1 *b*. The aerial at A is replaced by two similar aerials W and X whose outputs are added as in the vector diagram (figure 1 *c*) to give the output *E*(A). The aerial at B is replaced by two similar aerials Y and Z whose outputs are subtracted to give the output *E*(B) (as in figure 1 *d*). If all the aerials are identical, the outputs *E*(A) and *E*(B) are 90° out of phase. The relative amplitudes of *E*(A) and *E*(B) depend on the frequency received and on the aerial spacing but their phase difference is always 90° when the aerial array is symmetrically arranged about the incoming wave direction.

§ 3. ANALYSIS OF SYSTEM

Consider the aerial system at an orientation ϕ to the broadside position as in figure 1*f*. Let the distance $OW(=OX=OZ)$ be l and the wavelength along the ground of the incoming wave be λ. Put $k=2\pi l/\lambda$.

Let E be the amplitude of the voltage output from any one of the aerials. This will be very nearly the same for all aerials since the arms WZ, XY are some 30 feet long and thus the amplitude diversity effect is very small. Taking the centre point O as phase reference point and putting angle $XOW=YOZ=2\epsilon$, the signals from each aerial may then be written (omitting the $\exp(j\omega t)$ terms):

$$\left.\begin{aligned} E(\mathrm{W}) &= E\exp[jk\sin(\phi+\epsilon)] \\ E(\mathrm{X}) &= E\exp[jk\sin(\phi-\epsilon)] \\ E(\mathrm{Y}) &= E\exp[-jk\sin(\phi-\epsilon)] \\ E(\mathrm{Z}) &= E\exp[-jk\sin(\phi+\epsilon)] \end{aligned}\right\} \quad \ldots\ldots(1)$$

Thus

$$\left.\begin{aligned} E(\mathrm{A}) &= E(\mathrm{W})+E(\mathrm{X}) = 2E\cos(k\cos\phi\sin\epsilon)[\exp(jk\sin\phi\cos\epsilon)] \\ E(\mathrm{B}) &= E(\mathrm{Y})-E(\mathrm{Z}) = 2E\sin(k\cos\phi\sin\epsilon)[j\exp(-jk\sin\phi\cos\epsilon)] \end{aligned}\right\} \quad \ldots\ldots(2)$$

The amplitudes of the two output signals are

$$\operatorname{amp} E(\mathrm{A}) = 2E\cos(k\cos\phi\sin\epsilon),$$

$$\operatorname{amp} E(\mathrm{B}) = 2E\sin(k\cos\phi\sin\epsilon)$$

while the phase terms $\psi(\mathrm{A})$, $\psi(\mathrm{B})$ are in the square brackets. The amplitude of $E(\mathrm{B})$ goes to zero at $\phi=90°$ and also changes sign at that value. The amplitudes $E(\mathrm{A})$ and $E(\mathrm{B})$ have been plotted as a function of ϕ in figure 2*a* for a typical case, i.e. for $\epsilon=30°$ and $k=60°$, corresponding to the reception of a 30 m wave. In these curves, the amplitude of $E(\mathrm{B})$ has been drawn as being always positive, the change of sign at $\phi=90°$ being added to its phase angle. The phase angles $\psi(\mathrm{A})$ and $\psi(\mathrm{B})$ have also been plotted in figure 2*a*.

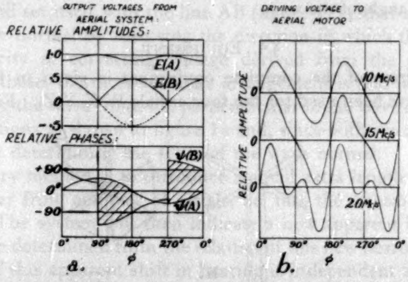


Figure 2. *a*, Relative amplitudes and phases of the signals from the aerials after addition and subtraction; *b*, relative driving voltage to servo-motor for different frequencies.

The controlling motor which aligns the aerial so that it is perpendicular to the wave normal depends for its operation on a determination of whether the phase difference $\psi(\mathrm{A})\sim\psi(\mathrm{B})$ is greater than or less than 90°. In the shaded regions of the phase plot in figure 2*a* this phase difference is greater than 90° and if the driving motor is arranged to increase ϕ in this circumstance the aerial array will take up one of the stable positions $\phi = 0^\circ$ or $\phi = 180^\circ$.

The controlling voltage for the d.c. driving motor is derived (in the way described by Whale 1954) from the signals $E(\mathrm{A})$ and $E(\mathrm{B})$ by combining them as follows:

$$\text{Let } E(\mathrm{A}) = A + ja, \quad E(\mathrm{B}) = B + jb.$$

Add $E(\mathrm{A})$ to $E(\mathrm{B})$ vectorially; the square of the magnitude of the resultant signal is given by

$$|E(\mathrm{A}) + E(\mathrm{B})|^2 = (A+B)^2 + (a+b)^2. \qquad \ldots\ldots(3)$$

Subtracting $E(\mathrm{B})$ from $E(\mathrm{A})$,

$$|E(\mathrm{A}) - E(\mathrm{B})|^2 = (A-B)^2 + (a-b)^2. \qquad \ldots\ldots(4)$$

The controlling voltage for the driving motor is proportional to the difference (3) − (4), i.e. the controlling voltage V is given by

$$\begin{aligned} V &\propto |E(\mathrm{A}) + E(\mathrm{B})|^2 - |E(\mathrm{A}) - E(\mathrm{B})|^2 \\ &\propto AB + ab \\ &= \mathscr{R}\{E(\mathrm{A})E^*(\mathrm{B})\} \qquad \ldots\ldots(5) \end{aligned}$$

where the asterisk denotes the complex conjugate.

From equations (2),

$$\mathscr{R}\{E(\mathrm{A})E^*(\mathrm{B})\} \propto \sin(2k\cos\phi\sin\epsilon)\sin(2k\sin\phi\cos\epsilon). \qquad \ldots\ldots(6)$$

This quantity is plotted in figure 2*b* for a system in which $\epsilon = 30^\circ$ and $l = 5$ m and for various values of λ. The aperture AB (in figure 1*f*) is then 8·7 m and the bearings obtained are unambiguous (except for the sense) for wavelengths greater than 17·3 m (frequencies less than 17·3 Mc/s). Graphs have been drawn for the frequencies 10, 15 and 20 Mc/s, the last one showing how ambiguities occur when the spacing AB exceeds half a wavelength. In practice, the upper frequency limit also depends on the vertical angle of incidence of the wave (increasing as the angle of incidence decreases) since the wavelength along the ground varies with the vertical angle of the wave.

§ 4. EQUIPMENT

A block diagram of the complete equipment is given in figure 3*a*. Two types of aerial have been used for the four aerials W X Y Z. For measurements

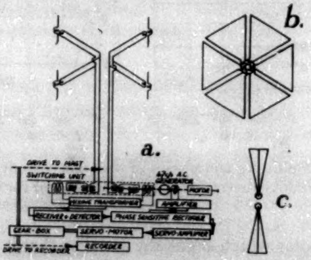


Figure 3. *a*, Block diagram of complete equipment; *b* and *c*, aerials for reception of horizontally and vertically polarized waves.

on the horizontally polarized component, horizontal loop aerials are used. In order that the phase of the output does not vary as the horizontal loop is rotated about a vertical axis, multiple loops of the form shown in figure 3 *b* are employed. These consist of six triangular sections connected in parallel and have the advantage over conventional loops that they possess low output impedance (desirable for operation over a wide band of frequencies) and also give zero phase shift as they are rotated about a vertical axis. For measurements on the vertically polarized component of the incoming wave, vertical dipoles are employed, the impedance being reduced somewhat by using fan-shaped aerials (figure 3 *c*). The band-width characteristics of the dipoles are inferior to those of the loop aerials but they have the advantage of giving a greater signal output.

The switching unit which carries out the vector additions and subtractions indicated in equations (3) and (4) contains two identical rotating-coil mechanisms similar to radiogoniometers with only one aerial coil. One of these mechanisms is associated with each aerial pair. One of them is fixed in position and serves only to introduce the same average phase change as the other which is motor driven and reverses the phase of the input signal at about 2800 r.p.m. (47 c/s). A small a.c. generator on the same shaft provides a 47 c/s reference signal for phase comparison purposes.

The radio-frequency output from the switching unit is amplified in a conventional receiver, detected, passed through a band-pass filter centred on 47 c/s, and this signal is then compared in phase with that from the a.c. generator in a conventional four-diode phase sensitive detector. A similar arrangement has been described previously (Whale 1954). The d.c. component of the output is then approximately proportional to $\mathscr{R}\{E(\mathrm{A})E^*(\mathrm{B})\}$, the approximation being best near the important zero points.

This d.c. output controls a servo-motor which drives the aerial system to the broadside-on position.

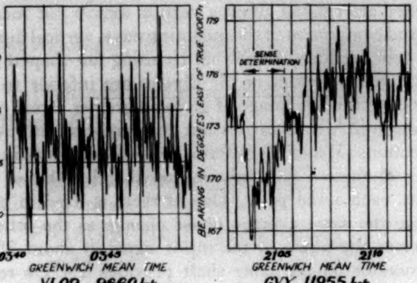
§ 5. DETERMINATION OF SENSE

With this system there are four symmetrical positions. It may be arranged that the system will set itself with the line AB (figure 1 *f*) either along or at right angles to the wave normal by reversing the direction in which the motor drives for a given polarity of correcting voltage derived from the phase-switching mechanism. Of these, the more sensitive arrangement is that in which the line AB sets itself along the wave front, i.e. perpendicular to the wave normal. There are two such positions as shown in figure 1 *e* and, since both are stable, provision must be made for determining the sense of the wave normal.

One satisfactory method is to introduce a small extra length of transmission line into the feeder from one pair of aerials, i.e. into the transmission line from either A or B. The system will then indicate a new apparent bearing and the sense can easily be determined from the relation of this new bearing to the former one. The size of this apparent shift in bearing is independent of the frequency but again somewhat dependent on the elevation angle if a fixed length of cable is introduced.

§ 6. BEARING RECORDS

A considerable number of records of bearing fluctuations has been obtained with this equipment, two representative samples being shown in figure 4. There are considerable and sudden variations in the character of such records. For stations at distances up to about 10 000 km the mean bearing is usually close to the great circle bearing, the fluctuations extending a mean angle of about 3° or so each side of this. For more distant stations the fluctuations are of the same order of magnitude but the mean bearing usually also shows slower variations which may be very large.



VLQ9 9660 kc/s
29 JUNE 1954
a.

GVY 11955 kc/s
28 SEPT. 1954
b.

Figure 4. Experimental records of fluctuation of the wave normal direction.

The amplitude of the fluctuations in the VLQ9 record is typical of many, of the records for stations distant about 2000–3000 km (VLQ9 Brisbane, 9660 kc/s is about 2300 km from Auckland). The other record is of a BBC short-wave station, GVY, on 11 955 kc/s.

Analysis of many records has shown that the time constant of the complete equipment is of the order of 5 to 10 seconds. However, the amplitude fluctuations (i.e. fading) of radio signals may sometimes have periods down to a second or so, indicating that the bearing fluctuations may also have periods as short as this. Improvements in the ratio of driving power to inertia of the aerials in the existing model will enable the time constant to be reduced and the study of more rapid bearing fluctuations to be undertaken.

§7. Actual and Measured Wave-Normal Directions

Some idea of the accuracy with which this instrument fulfils its function of measuring the fluctuations in the direction of the horizontal component of the wave normal of a radio wave may be gained from the following. Wave normal fluctuations similar to those actually observed could be produced if the incoming wave consisted of a series of plane waves of varying phase clustered around a few main directions and usually with wave normals within ± 5° of these average directions. To illustrate the behaviour of this direction finder we consider a simplified model in which the incoming wave is made up of only two component plane waves. The effects arising from the interference of two waves are well known in direction finding (Heiligtag 1923) and radar (Kerr 1951).

Let these two waves, both of wavelength along the ground equal to $2\pi/K$, have wave functions in the horizontal (xy) plane given by

$$\exp jKx \qquad \text{......(7)}$$

and

$$\alpha \exp j\{K(x\cos\theta - y\sin\theta) + \xi\} \qquad \text{......(8)}$$

where (7) represents a plane wave of unit amplitude travelling in the positive x direction, and (8) represents a plane wave of amplitude α travelling in the direction $-\theta$ (measured from the direction of the positive x axis) and with its phase advanced ξ with respect to wave (7) when measured at the point $O(x=0, y=0)$. The angle ψ between the wave normal of the resulting wave (measured at O) and the direction of the positive x axis is given by

$$\tan\psi = -\frac{\alpha\sin\theta(\alpha+\cos\xi)}{(1+\alpha\cos\xi)+\alpha\cos\theta(\alpha+\cos\xi)}. \qquad \ldots\ldots(9)$$

The shape of the resultant wave front may be obtained from equation (9) by an integration since $\tan\psi = -dy/dx$ and ξ, the phase difference between the two component waves, is a function of x and y. A typical wave front pattern is shown in figure 6*a* where the wave fronts are moving in the direction LM with a phase velocity $c/\cos\frac{1}{2}\theta$ (θ being the angle between the two components). If ξ is altered the whole pattern will be shifted sideways across the page so that a direction finder at a fixed point will indicate a different wave normal direction.

The wave normal direction indicated by the direction finder has been shown in equation (5) to be given by the position of the zeros in the expression for $\mathscr{R}\{E(\mathrm{A})E^*(\mathrm{B})\}$.

In the case of two component waves and using the same symbols as in expression (1):

$$\begin{aligned}\mathscr{R}\{E(\mathrm{A})E^*(\mathrm{B})\} &= \sin(2k\sin\epsilon\cos\phi)\sin(2k\cos\epsilon\sin\phi)\\ &+\alpha^2\sin\{2k\sin\epsilon\cos(\phi-\theta)\}\sin\{2k\cos\epsilon\sin(\phi-\theta)\}\\ &+2\alpha\cos\xi\sin[k\sin\epsilon\{\cos\phi+\cos(\phi-\theta)\}]\sin[k\cos\epsilon\{\sin\phi+\sin(\phi-\theta)\}]\\ &-2\alpha\sin\xi\sin[k\sin\epsilon\{\cos\phi-\cos(\phi-\theta)\}]\cos[k\cos\epsilon\{\sin\phi+\sin(\phi-\theta)\}].\end{aligned} \qquad \ldots\ldots(10)$$

If the aperture of the direction finder is very small (i.e. $k=2\pi l/\lambda$ is small) and if θ and ϕ are small, the indicated wave normal direction ϕ is equal to the true wave normal direction ψ given by equation (9). In general, k is not negligibly small. For example, with an arm-length of 5 metres and an operating wavelength of 30 metres, $k=\pi/3$. The actual wave normal direction and the indicated direction have been calculated for the case where the angle θ between the two constituent 30 metre waves is 10°, the variation in wave normal direction being obtained by letting ξ vary steadily from 0° to 360°. Examples are shown in figure 5 for the cases where α (the ratio between the amplitudes of the waves) is equal to 0·9 and 1·11, i.e. there is about ± 1 db difference in level. For larger differences in level the excursions of wave normal direction are not so great and the direction finder follows the true wave normal direction more accurately.

The way in which the indicated direction depends on the aperture of the direction finder can be seen from a comparison of the scale of the irregularities in the wave front with the size of the direction finder. Actual wave fronts are plotted in figure 6*a* for the same two waves considered above. The constituent waves are indicated in figure 6*b*. The full lines are for the case of $\alpha=0{\cdot}9$ and the broken lines for the case of $\alpha=1{\cdot}11$. It will be seen that the mean wave front is determined by the stronger of the two waves but that the most common wave normal direction is intermediate between the wave normal directions of the two component waves.

The differences between the indicated and true wave normal directions drawn in figure 5 arise in two ways. Firstly, the true wave normal direction at any point is the normal to a tangent to the wave front at that point. If the curvature of wave front is appreciable, any method which derives the tangent from measurements of phase at two non-coincident points will introduce error, i.e. the measured direction is perpendicular to a chord of the wave front rather than to a tangent. This amounts to a 'smoothing' of the wave front over the aperture of the direction finder. This smoothing accounts for the main differences between the curves in figure 5. Secondly, the diversity effect is not completely negligible in the case considered. There are troughs of resultant field strength near the places where the maxima in the wave normal deviation occur so that the aerials may pick up signals of different amplitude. This effect introduces a slight asymmetry about the $\xi = 180°$ line in the curves for indicated direction

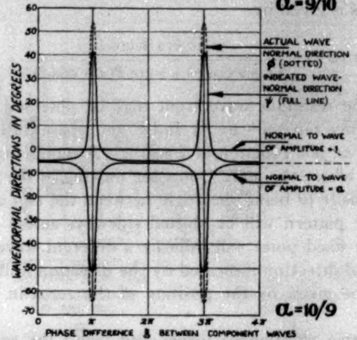
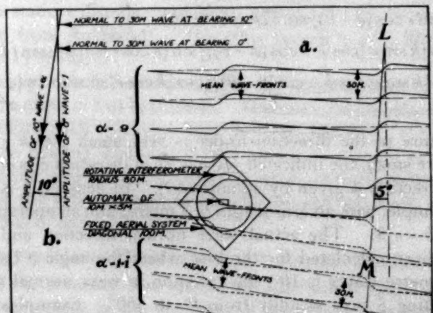


Figure 5. Actual and indicated wave normal direction when the incoming signal consists of two waves in directions differing by 10° and with strengths differing by ± 1 dB.



NORMAL TO 30M WAVE AT BEARING 10°
NORMAL TO 30M WAVE AT BEARING 0°
a.
L
MEAN WAVE-FRONTS
30M
α=9
AMPLITUDE OF 10° WAVE =α
AMPLITUDE OF 0° WAVE =1
10°
ROTATING INTERFEROMETER RADIUS 80M
AUTOMATIC DF 10M × 3M
FIXED AERIAL SYSTEM DIAGONAL 100M
5°
b.
α=1·1
MEAN WAVE-FRONTS
M
30M

Figure 6. *a*, Wave front pattern for two unequal waves with 10° difference in bearing ; *b*, the wave normal directions of the constituent waves.

but this is not apparent on the scale to which the curves in figure 5 are drawn. It arises from the last term in equation (10).

The sizes of some representative direction finding systems are shown in figure 6*a*. The automatic direction finder described here is approximately 10 m by 5 m; spaced aerial systems with the aerials fixed at the corners of a square with diagonals of 100 m are in use; the rotating radio interferometer has a radius of 60 m. Since the wave normal direction indicated by a direction finder is averaged over approximately the area occupied by the aerial array, the systems with wide spacing will not generally show such large fluctuations of wave normal direction as will systems with small apertures.

The wave fronts drawn in figure 6*a* are for two component waves in directions differing by 10° corresponding to an observed case (Ross 1955). When the spread of component waves is less than this (as is usually the case when one isolated ionospheric wave has been selected for reception) the scale of the pattern relative to the size of the direction finder is increased so that it is more likely that the true wave normal direction will be measured by the systems with large aperture.

If the field-pattern in figure 6*a* were stationary, a two-aerial direction finder would indicate a steady direction of the wave normal. In order to resolve the two component waves the two aerials of the direction finder would need to be about λ/S apart (S = sine of the angle between the two waves) and the space between these two extreme aerials would need to be filled by other aerials spaced less than approximately λ apart in order that ambiguities would be avoided. Typical curves illustrating these now well-known requirements were given by Friis and Feldman (1937). Measurements could be made simultaneously on the signals from all the aerials in this array but the same results could be obtained by making measurements on each pair of adjacent aerials (spaced about λ apart) in turn, this being equivalent to making measurements with a pair of aerials of fixed spacing which is moved (or scanned) across the total aperture of the direction finder.

We are concerned with a different problem from the above stationary-field case since the whole field pattern moves across the stationary direction finder as the phase differences between the component waves change. We are then, in effect, 'scanning' the field with the direction finder. The way in which this scanning occurs can be calculated for some simple field distributions. For example, in the idealized two-component case, the direction finder moves steadily across the field as the phase difference between the two components changes steadily so that the total effective aperture of the direction finder is very large indeed, and the closer the aerial spacing the less the chance of ambiguities in the measurements, i.e. the finer the structure of the pattern which can be measured.

When the component waves are randomly related to each other the situation is much more complicated and only the statistical properties of the distribution of incoming waves can be observed. This case has been treated by Bramley (1951).

An example of the large bearing fluctuations which are associated with the presence of two main components in the incoming wave is shown in figure 7. The relative amplitude of the signal is also shown in this diagram. The amplitude of the resultant wave is proportional to $\sqrt{(1+\alpha^2+2\alpha\cos\xi)}$ and will thus be a minimum when $\cos\xi = -1$, i.e. the two waves are in anti-phase. This is also the condition for the large excursions of the bearing (wave normal direction) from the average. It can be seen from the experimental curves that the amplitude minima occur at the same time as the large bearing excursions, whether these are towards higher or lower bearings. The reversal of the direction of the bearing excursion at about 0948 (compared with those preceding and following it) is probably due to a change in the magnitude of α from a value less than 1 to a value greater than 1.

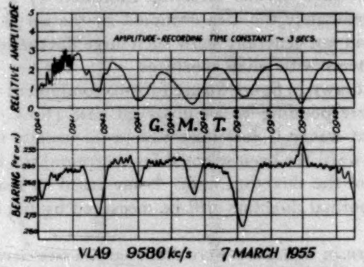


Figure 7. Experimental record of amplitude and bearing fluctuations showing the field strength minima associated with large bearing excursions.

Records (as in figure 7) of fluctuations of both amplitude and wave normal direction provide a guide to the acceptability of the results. Since both quantities are essentially interference phenomena, the rapidity of the amplitude fluctuations is an indication of the rapidity to be expected in the bearing fluctuations. The former can be recorded with very small time constants and is not limited by the inertia of the aerial system. From comparisons of this nature it has been found that the present equipment is adequate for recording bearing fluctuations associated with all but the most rapid types of amplitude fading.

ACKNOWLEDGMENTS

This work was financed by a University of New Zealand Research Grant. One of us (W. J. R.) held a University of New Zealand Research Fellowship during this period.

The Seagrove Radio Research Station is operated by the Department of Physics, Auckland University College, and receives a grant from the New Zealand Department of Scientific and Industrial Research.

REFERENCES

Bramley, E. N., 1951, *Proc. Instn Elect. Engrs*, Pt III, **98**, 19.
Bramley, E. N., and Ross, W., 1951, *Proc. Roy. Soc.* A, **207**, 251.
Friis, H. T., and Feldman, C. B., 1937, *Bell. Syst. Tech. J.*, **16**, 337.
Heiligtag, T., 1923, *Jahrb. d. Draht.*, **21**, 77.
Kerr, D. E., 1951, *Propagation of Short Radio Waves*, Radiation Laboratory Series, **13**, 441 (New York and London : McGraw-Hill).
Miya, K., and Ishikawa, M., 1955, *Patent News*, No. 29–3418 (Direct-vision type direction-finder), Radio Management Bureau, G.P.O., Japan.
Ross, W. J., 1955, *Ph.D. Thesis*, Auckland University College, p. 54.
Whale, H. A., 1954, *Proc. Phys. Soc.* B, **67**, 553.

The Radar Measurement of Low Angles of Elevation

By J. S. HEY and S. J. PARSONS†
Ministry of Supply, Radar Research Establishment, Malvern, Worcs.

MS. received 6th October 1955

Abstract. The errors which occur in the measurement of low angles of elevation by centimetric radar on a flat ground site are estimated theoretically and compared with experimental measurements. It is demonstrated that a very considerable improvement in angular accuracy is obtained by erecting a screen to intercept the ground-reflected wave. The diffraction effects due to the screen are discussed.

§ 1. Introduction

For any given aerial aperture of a radar equipment it is possible to obtain a narrower beam and consequently a higher accuracy of angular measurement by using a shorter wavelength. Centimetric wavelengths are consequently preferable to wavelengths of several metres when accurate angular measurement is required. On the usual type of ground site there is, however, a minimum angle of elevation, of the order of the beam width, for which accurate elevation measurement is possible. At lower angles of elevation serious errors may occur since, in addition to the radiation received directly from the elevated object under observation, an appreciable amount enters the receiver aerial by reflection from the ground.

A previous report by Hey, Parsons and Jackson (1947) described the effect of ground reflection on the maximum detection range of centimetric radar equipments directed at low angles of elevation. An account will now be given of a theoretical and experimental research to determine the effect of ground reflection on angular accuracy. The interception of the ground-reflected waves by means of an artificial screen is investigated; in this case, the primary factor affecting accuracy arises from diffraction at the screen edge.

§ 2. The Radar Equipment

The radar equipment operated on a wavelength of 10·7 cm. The aerial system consisted of separate transmitter and receiver dipoles with parabolic reflectors of 1·22 m diameter. These aerials were mounted adjacently with their centres at a height of 3·6 m and could be traversed together in elevation or bearing as required.

Angular measurement by means of this radar (the GL3 Army Radar) has been described in detail by Beeching (1946). To measure the angle of elevation of a radar target‡ the receiver output is observed when the receiver dipole is alternately deflected equal amounts above and below the focus of the parabolic reflector. The directional sensitivity pattern of the receiver aerial is shown in figure 1, curve *a*. The movement of the receiver dipole swings the aerial beam alternately 2·4° above and 2·4° below the aerial axis. The angle of elevation of the target is indicated by the direction of the aerial axis when adjusted so that the receiver output is the same for both deflected positions of the aerial beam.

---

† Now at Mersey Cable Works Ltd.

‡ In this paper, we refer to the object under observation by the radar as the radar target. The ' angle of elevation ' is often referred to as ' the elevation '.

The radar target used in the investigation was a papier-mâché sphere, of 0·6 m diameter, sprayed with zinc and suspended from a lightweight balloon tethered by a nylon cord. At high elevations of the target a negligible amount of the radiation reflected from the ground is received by the aerial and the elevation of the target can be determined to an angular accuracy of a few minutes of arc. At elevations within a few degrees of the ground the directional sensitivity of the receiver aerial is wide enough to accept a considerable amount of radiation reflected from the ground, and serious errors may result. A theoretical assessment of possible errors when the radar is sited on flat ground is given in the following section.

§ 3. THEORETICAL ESTIMATION OF ERRORS OVER FLAT GROUND

The beam of aerial sensitivity in the vertical plane in free space can be described by a function $f(\theta)$ which is the received amplitude from a plane wave at angle θ from the direction of the maximum. For elevation measurement the receiver aerial beam is alternately deflected an angle $\pm\delta$ in elevation with respect to the aerial axis. Hence if the target is at an angle of elevation α and the aerial axis at elevation β, the values of $f(\theta)$ in the two deflected positions of the beam correspond to $\theta=\beta-\alpha\pm\delta$. Since $f(\theta)$ is sensibly symmetrical with respect to positive and negative values of θ, and the same for both deflected positions, the values of $f(\theta)$ are equal when $\beta-\alpha=0°$. The elevation of the target is then indicated by the direction of the aerial axis.

In practice, when the radar equipment is observing at low elevations on a ground site it is necessary also to take into account the ground-reflected wave. The errors in elevation measurement to be expected may be calculated in certain idealized cases. We shall assume that the ground is flat and has a uniform reflection coefficient. There is then an image target in a direction at an angle $-(\alpha+\beta)$ from the aerial axis.

Let r be the reflection coefficient, λ the wavelength, h the height of the centre of the aerial above ground and ϕ the phase difference between direct and reflected waves $=(4\pi h/\lambda)\sin\alpha$.

The received amplitude in the two deflection positions of the beam will be the same when

$$|f(\theta)_{\beta-\alpha+\delta}+re^{i\phi}f(\theta)_{\beta+\alpha+\delta}|=|f(\theta)_{\beta-\alpha-\delta}+re^{i\phi}f(\theta)_{\beta+\alpha-\delta}|.$$

Thus if r is known the value of β satisfying the above equation may be determined graphically for any given value of α. The theoretical errors in elevation measurement are then equal to $\beta-\alpha$.

If the ground were a perfectly reflecting surface, then $r=-1$; all other types of ground give $r=-1$ for $\alpha=0°$. The investigation of Hey, Parsons and Jackson (1947) indicated that for elevations up to about 8° at $\lambda=10{\cdot}7$ cm the reflection coefficient maintains a comparatively high numerical value on natural ground sites. The elevation errors $\beta-\alpha$ for the theoretical flat-ground site have been determined for values of $r=-1$, $r=-0{\cdot}8$, and $r=-0{\cdot}5$, for the target elevations from 0° to 6°, and the results are shown in figure 2.

These results indicate not only that very large errors may occur, but also that there are often three positions of the aerial axis which satisfy the condition of equal received amplitude for the two deflected positions of the beam. One of these positions, indicated by the broken line in figure 2, may be distinguished by means of a sense rule with respect to a small change in elevation of the aerial.

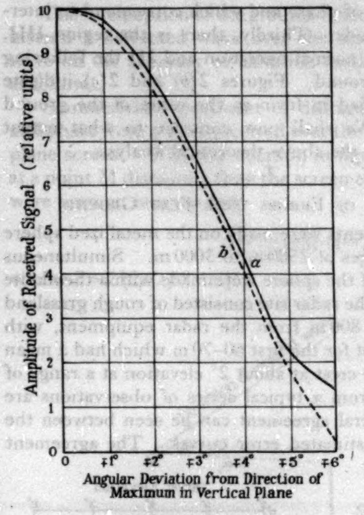


Figure 1. Directional sensitivity pattern of receiver aerial: (*a*) experimental, (*b*) theoretical.

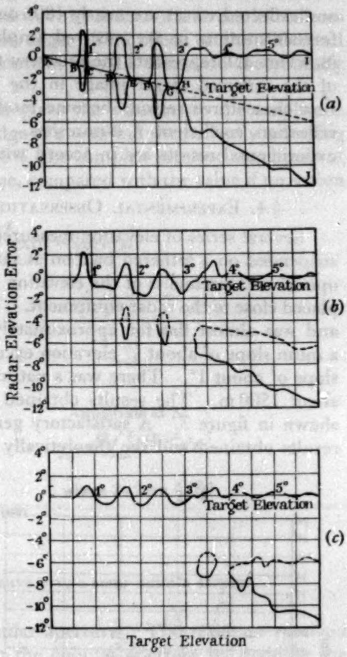


Figure 2. Theoretical errors over flat ground. Reflection coefficient: (*a*) $-1{\cdot}0$, (*b*) $-0{\cdot}8$, (*c*) $-0{\cdot}5$.

The sense rule which applies when ground reflection can be neglected is as follows. When the received amplitude in the upper deflected position of the beam is less than that in the lower position then this indicates that an increase in aerial elevation is required in order to equalize the two amplitudes. When the aerial is near the position indicated by the broken line in figure 2 a movement according to the usual sense rule would accentuate the difference between the received amplitudes. Nevertheless, it would be a source of confusion in radar operation.

Consider now figure 2 (*a*) corresponding to $r = -1$. There are three distinct types of region of elevation error. First, the region typified by AB, CD, EF, or

GH, in which the aerial axis remains at zero elevation. These regions correspond to the interference maxima in received amplitude (see Hey, Parsons and Jackson 1947, p. 849). The maxima are a consequence of equal received amplitudes of direct and ground-reflected waves with little phase difference between them. Secondly, in the regions BC, DE, FG, there are two positions both giving the correct sense rule. The error changes rapidly in these regions in which direct and reflected waves are nearly 180° out of phase and which correspond to interference minima in the received amplitude. Thirdly, there is the region IHJ, in which HI represents the recovery to normal operation and HJ the following of the reflected target image in the ground. Figures 2(*b*) and 2(*c*) indicate how these three regions become modified in form as the value of the ground reflection coefficient *r* diminishes. We shall now consider to what extent experimental results are in accord with the above theoretical analysis.

§ 4. EXPERIMENTAL OBSERVATION OF ERRORS OVER FLAT GROUND

Several series of elevation measurements were made on the metallized sphere suspended on a tethered balloon at ranges of 2500 m to 3000 m. Simultaneous optical measurements of the elevation of the sphere were made with a theodolite placed close to the radar equipment. The radar site consisted of rough grassland and was almost flat for approximately 800 m from the radar equipment with a mean slope of about ½° elevation except for the first 60–70 m which had a mean slope of about 1°. There was a natural crest at about 2° elevation at a range of about 1500 m. The results obtained from a typical series of observations are shown in figure 3. A satisfactory general agreement can be seen between the results obtained and the theoretically estimated error curves. The agreement

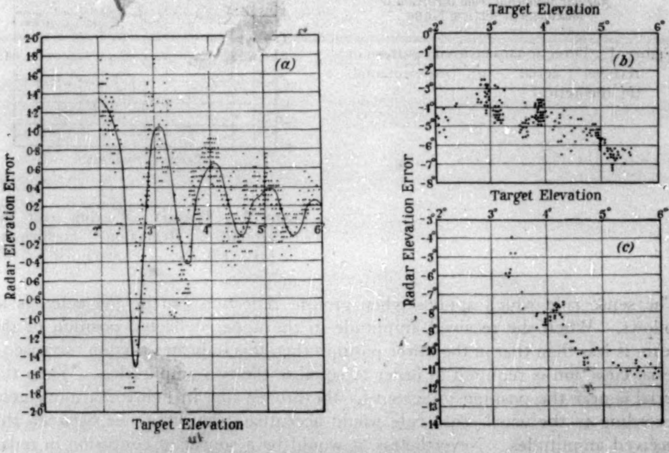


Figure 3. Observed errors over ground (mean slope of ground ½° to 1°).
(*a*) Continuous line indicates mean error.
(*b*) Errors due to following in reversed sense rule.
(*c*) Errors due to following image in ground.

appears to be best for $r = 0{\cdot}8$. Thus the results in figure 3 (*a*) are seen to correspond approximately to the upper error curve in figure 2 (*b*) when due allowance is made for the slope of the ground. Similarly, the results in figure 3 (*b*) and 3 (*c*) correspond respectively to the broken and lower curves of figure 2 (*b*).

§ 5. THEORETICAL ESTIMATION OF ERRORS OVER AN ARTIFICIAL SCREEN

The interception of the ground-reflected wave by means of a screen introduces instead a diffracted wave over the screen edge. The effect of the diffracted wave on the accuracy of elevation measurement may be estimated theoretically as follows.

Sommerfeld's treatment of diffraction gives the solution to the amplitude and phase from a wave received over the edge of a semi-infinite perfectly reflecting plane screen. With the notation in figure 4, the complex received amplitude A at a point M distance ρ from the screen edge, compared with the value if no screen were present, is given by[†]

$$A = \frac{1}{\sqrt{2}} \exp\left(\tfrac{1}{4}i\pi\right) \int_{-\infty}^{Z} \exp\left(-\tfrac{1}{2}i\pi\nu^2\right) d\nu$$

where $Z = 2\,(2\rho/\lambda)^{1/2} \sin \tfrac{1}{2}\delta$.

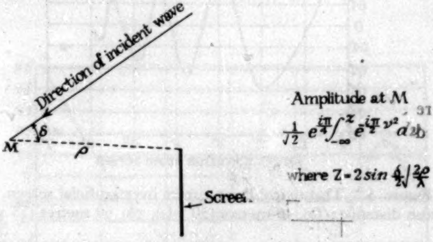

Direction of incident wave
δ
M
ρ
Screen.
Amplitude at M
$\frac{1}{\sqrt{2}} e^{\frac{i\pi}{4}} \int_{-\infty}^{z} e^{-\frac{i\pi}{2}\nu^2} d\nu$
where $Z = 2 \sin \frac{\delta}{2} \sqrt{\frac{2\rho}{\lambda}}$

Figure 4. Diffraction of plane wave over a semi-infinite screen.

The actual aerial receives over a circular aperture. The resultant received amplitude was derived from the wave at the circular aperture by dividing the aperture into horizontal strips. Let S be the area of a strip expressed as a fraction of the whole aperture. Then the contribution to the total received amplitude from the strip is SA where A has the mean value for the strip. The resultant amplitude, compared with the maximum obtained in free space, is then given by the complex summation of the contributions from all the strips.

The summation may be carried out for any given elevation of the incident wave for various inclinations of the aerial aperture. It is thus possible to determine graphically a position of the aerial axis such that the magnitude of the received amplitudes is the same for deflections of ± δ in elevation. This direction will be the apparent elevation of the target as indicated by the radar. Error curves calculated by the above method for the equipment used are shown in figure 5 for distances of 46 m (50 yards) and 69 m (75 yards) from the screen.

[†] The derivation of this equation may be obtained, for example, from Baker, B. B., and Copson, E. T., (1950), *The Mathematical Theory of Huygens' Principle* (Oxford: Clarendon Press), pages 142–4; it should be noted that the second term in their expression for the diffracted field strength becomes negligible when as, in the above case, ρ/λ is large and the angles between the normal to the screen and the directions of the incident and diffracted waves are small.

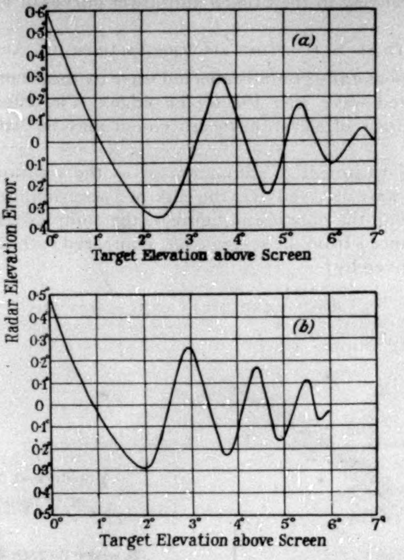


Figure 5. Theoretical error curves over artificial screen. Screen distance : (*a*) 46 metres (50 yds), (*b*) 69 metres (75 yds).

Although the technical method by which the deflection is achieved is by displacement of the dipole with respect to the focus of the parabolic reflector, this produces a phase change across the aperture, which is equivalent to changing the inclination of the aperture as considered in the theoretical treatment. The assumption of uniform sensitivity to received amplitude over the aerial aperture is only approximately correct. The comparison in figure 1, curve *b* of the theoretical amplitude pattern for free space with the experimental one indicates a satisfactory approximate agreement. An improved agreement could have been obtained by introducing weighting factors for the strip areas, but this was not considered necessary since the calculated errors of elevation measurement are comparatively insensitive to small changes in the aerial beam width.

A further second-order effect in the practical case may be introduced by ground reflection of the diffracted waves. When the screen distance is only several times the height of the radar this diffracted wave reflected from the ground between the set and the screen is of very small amplitude and becomes infinitesimal when the aerial directivity is taken into account. A ground-reflected wave on the far side of the screen could be more important at angles of elevation approaching zero. Over land it is in practice unusual to have flat ground sufficiently extensive and the natural horizon sufficiently low for the ground-reflected wave to develop at elevations approaching zero. In the experimental conditions here described the variation of distant ground contour and the occurrence of a natural crest justified the neglect of diffracted waves reaching the aerial after the ground reflection.

§ 6. EXPERIMENTAL OBSERVATION OF ERRORS OVER AN ARTIFICIAL SCREEN

An artificial screen of corrugated iron was erected with its top edge horizontal and with a vertical extent from the ground to an elevation of about 1·8° from the centre of the receiver aerial so as just to obscure a natural crest at a distance of 1500 m. Several series of observations on the spherical reflector flown at ranges of 2000–3000 m were made for distances of 46 m (50 yards) and 69 m (75 yards) between the radar equipment and the screen. For each series the difference was observed between the elevation of the reflector as indicated by the radar and as measured optically by means of a theodolite. The results are shown in figure 6 which indicates excellent agreement between the errors assessed theoretically and those measured experimentally. It will be seen that these errors are considerably smaller in magnitude than those which occur over flat ground with no screen, and that they are free from ambiguities.

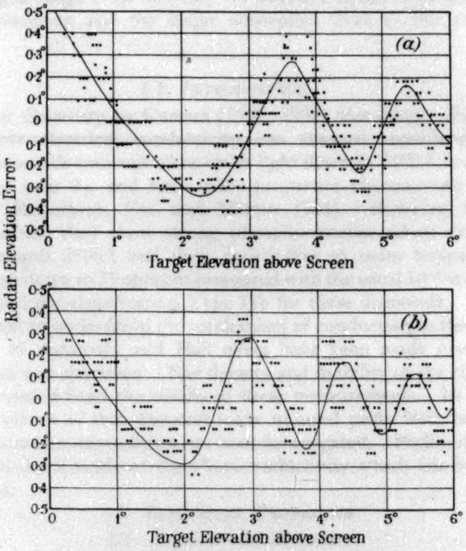


Figure 6. Observed errors over artificial screen.
Screen distance : (*a*) 46 metres (50 yds), (*b*) 69 metres (75 yds).
(The theoretical curves are shown in continuous line for comparison.)

§ 7. CONCLUSION

The above theoretical and experimental analysis has shown that the large errors and ambiguities which occur in radar elevation measurement at low angles of elevation due to ground reflection, can be very much reduced by the introduction of an artificial screen at an appropriate distance to intercept the ground-reflected wave. Diffraction at the screen edge introduces a small error, but this is free from ambiguity and an appropriate correction may be applied.

The authors are indebted to F. Jackson for his help in both the experimental and computational work.

REFERENCES

BEECHING, G. H., 1946, *J. Instn Elect. Engrs*, Pt. IIIA, **93,** 519.
HEY, J. S., PARSONS, S. J., and JACKSON F., 1947, *Proc. Phys. Soc.*, **59,** 847.

Electrical and Optical Properties of a Semiconducting Diamond

By I. G. AUSTIN and R. WOLFE
Research Laboratories of The General Electric Company, Ltd, Wembley, Middlesex

MS. received 21st September 1955 and in final form 6th December 1955

Abstract. The electrical resistivity and Hall coefficient of a semiconducting diamond (Type IIb) have been measured between -100°c and 600°c using standard d.c. methods. The diamond behaves like a normal p-type semiconductor. At 20°c its resistivity is 270 ohm cm and the Hall mobility of the holes is $1550 \pm 150\,\mathrm{cm^2\,volt^{-1}\,sec^{-1}}$. Near this temperature the mobility varies approximately as $T^{-3/2}$. The activation energy of the acceptors is 0·38 ev. It is suggested that these acceptors are impurity atoms and that the total concentration of impurities is smaller in Type IIb diamonds than in other diamonds.

The infra-red absorption spectrum of this diamond is typical of Type II diamonds but it shows an additional prominent peak at $3{\cdot}57\,\mu$, smaller peaks at $4{\cdot}1\,\mu$, $3{\cdot}4\,\mu$ and $2{\cdot}5\,\mu$, and a continuum of absorption in the range 1 to $2{\cdot}5\,\mu$. This extra absorption is reversibly temperature dependent. The intensity decreases with increasing temperature and the peaks are undetectable above 300°c.

At -155°c, photocurrents of the order of 10^{-9} amp have been detected in the wavelength range $0{\cdot}9\,\mu$ to $3{\cdot}6\,\mu$. A tentative model is proposed to relate this photoconductivity and the major absorption peak to the mechanism of conduction.

---

§ 1. Introduction

The recent discovery by Custers (1952, 1954) that certain diamonds have appreciable electrical conductivity has aroused considerable interest. These diamonds transmit ultra-violet light down to 2300 Å, are transparent in the infra-red near $8\,\mu$ and have other properties characteristic of Type II diamonds (cf. Robertson, Fox and Martin 1934). However, unlike other Type II diamonds, they show strong phosphorescence when irradiated with light of wavelength 2500 Å and their resistivities at room temperature range from 10^8 ohm cm down to 25 ohm cm compared with the usual 10^{14} or 10^{16} ohm cm. Custers proposed the classification Type IIb for these diamonds.

In an attempt to understand the mechanism of conduction in these diamonds, measurements of resistivity and Hall effect have been made over a range of temperatures on one specimen. The density and mobility of the charge carriers have been calculated from the results of these measurements. In the infra-red absorption spectrum of this diamond some unusual peaks have been observed and their temperature dependence has been investigated. Preliminary measurements have also been made of the photoconductivity which has been observed in the infra-red.

§ 2. Electrical Properties

2.1. *Experimental Procedure*

The diamond is a 'natural blue' rectangular block of dimensions $2{\cdot}99\,\mathrm{mm} \times 1{\cdot}53\,\mathrm{mm} \times 1{\cdot}52\,\mathrm{mm}$. With an indium contact pressed on to one end and a sharp tungsten probe on the other, a p-type rectifying characteristic was observed on an oscilloscope. For most of the experiments electrical contact was made by pressing graphite blocks against the ends of the diamond and graphite or copper probes on its edges. These contacts were variable and non-ohmic but the contact resistances were much smaller than those obtained when probes were placed on the smooth faces of the crystal.

In preliminary end-to-end resistance measurements, it was found that the resistance fell as the diamond was heated and rose again when the temperature decreased, but when the temperature was returned to 20°c the resistance was as much as ten times lower than its original room temperature value (cf. the 'hysteresis' observed by Custers (1955)). This reduction was found to be caused by changes in the contact barriers and not in the bulk of the diamond. When resistivity measurements were made using potential probes the results were consistent and repeatable.

The resistivity and Hall effect measurements were made with a standard potentiometer and galvanometer circuit. As the temperature was lowered the increasing impedance of the diamond limited the sensitivity of this apparatus but the larger voltages could still be measured with errors not larger than 5% down to -100°c. To prevent oxidation the high temperature measurements were made with the crystal in an argon atmosphere. To avoid the formation of a graphite layer on the surface no measurements were made above 600°c. The electromagnet used for the Hall effect measurements produced a field of 2130 oersteds. No magnetoresistance effect was detected with fields up to 4000 oersteds. The low temperature Hall effect measurements were repeated using a vibrating reed electrometer in place of the galvanometer as a null detector. The results obtained by the two methods were identical.

Temperatures were measured with a fine thermocouple which was pressed against the diamond. To avoid errors due to temperature differences between the specimen and its surroundings, the current through the specimen was kept small.

2.2. *Results*

At 20°c the resistivity of this diamond is 270 ohm cm. The positive Hall coefficient of $4{\cdot}2\times10^5$ cm^3 coulomb^{-1} at this temperature indicates that the current carriers are holes in agreement with the p-type rectifying characteristic. Since the diamond is only twice as long as it is wide, the Hall voltages were smaller than those which would be obtained in a specimen of length-to-width ratio greater than four (Isenberg, Russell and Greene 1948, Dunlap 1950). A correction factor of 1/0·93 was therefore applied in calculating the Hall coefficient from the measured Hall voltages. The density of holes P was calculated from the usual semiconductor equation

$$R = 3\pi/8P\boldsymbol{ec} \qquad \ldots\ldots(1)$$

where R is the Hall coefficient, $\boldsymbol{e}$ is the electronic charge and $\boldsymbol{c}$ is the velocity of light. At room temperature the result is $P = 1{\cdot}8\times10^{13}$ cm^{-3}. The Hall mobility μ_H defined by $\mu_H = R/\rho$ is 1550 ± 150 cm^2 volt^{-1} sec^{-1} at 20°c.

The resistivity over the range of temperature from -100°c to 600°c is shown in figure 1. A resistance minimum of 1·5 ohm cm is reached at about 500°c. Below -50°c the plot of $k \log \rho$ against $1/T$ where T is the absolute temperature is a straight line with a slope corresponding to an energy of 0·38 ev.

In figure 2 the Hall coefficient and the density of holes are plotted as a function of reciprocal temperature. The shape-correction factor has been included in calculating these values from the measured Hall voltages, and equation (1) has been used over the whole temperature range. The maximum density has not been reached at 600°c but the shape of the curve indicates a saturation value of 2×10^{16} holes per cm³.

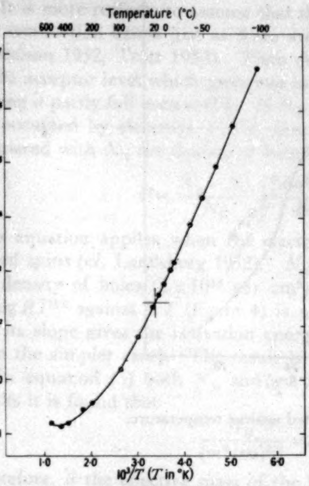


Figure 1. Resistivity plotted against temperature.

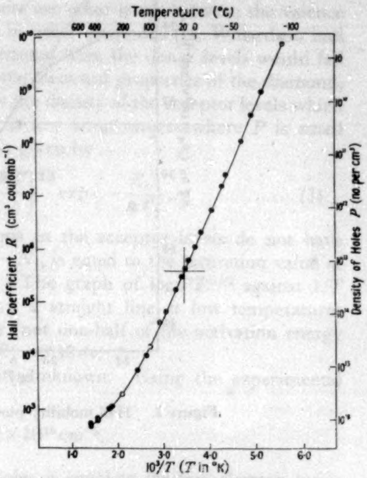


Figure 2. Hall coefficient and density of holes plotted against temperature.

The variation with temperature of the Hall mobility calculated from the curves in figures 1 and 2 is shown in figure 3. Near room temperature μ_H varies as $T^{-3/2}$ as is predicted by simple deformation potential theory (Bardeen and Shockley 1950). At low temperatures the mobility falls below the $T^{-3/2}$ line probably because impurity scattering becomes more important. The high temperature variation may be associated with scattering by the higher concentration of ionized acceptors but the possibility of systematic errors up to 10% of the measured values cannot be excluded.

2.3. *Discussion of Electrical Results*

The diamond behaves like a normal p-type impurity semiconductor. If we consider the simplest model which could account for its properties—a single acceptor level of density $N_A = 2 \times 10^{16}$ per cm³ just above the valence band—and if we assume that all of these acceptors would be empty of electrons at 0°K, then at low temperatures the density of holes would be given by

$$P = N_A^{1/2}(2\pi m^* \boldsymbol{k}T/\boldsymbol{h}^2)^{3/4} \exp(-\tfrac{1}{2}E_A/\boldsymbol{k}T) \quad \ldots\ldots(2)$$

where E_A is the activation energy of the acceptors, and m^* is the effective mass of the holes. The graph of $\log PT^{-3/4}$ against $1/T$ should therefore be a straight line, as is observed with the low temperature experimental points. (This graph is not shown but it is similar to the graphs in figures 2 and 4 since the power of T which is used to multiply P has very little effect on the slope of the straight line portion.) At any point on this line, the slope of which gives $\tfrac{1}{2}E_A$, everything in

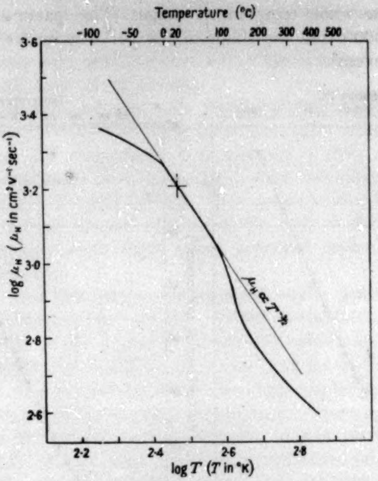


Figure 3. Hall mobility plotted against temperature.

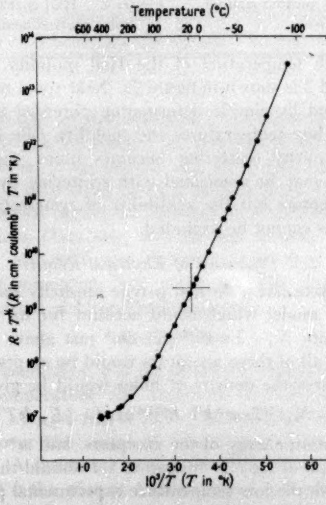


Figure 4. Hall coefficient × $T^{3/2}$ plotted against temperature.

this equation is known except m^*. The calculated result is $m^*/m = 500$ (m is the mass of the free electron). This result seems most unlikely when it is compared with the figures found for germanium and silicon in which $m^*/m < 1$ (Dresselhaus, Kip and Kittel 1955). It is therefore most probable that this model is too simple.

It is more realistic to assume that there are other levels between the valence and conduction bands just as there are in other diamonds (cf. Willardson and Danielson 1952, Trott 1953). Then electrons from the donor levels would fall to the acceptor level which gives rise to the electrical properties of the diamond, leaving it partly full even at 0°K. If N_D is the density of the acceptor levels which are occupied by electrons at 0°K, then at low temperatures where P is small compared with N_D the density of holes is given by

$$P = \frac{N_A - N_D}{N_D}\left\{\frac{2\pi m^* kT}{h^2}\right\}^{3/2} \exp\left\{-\frac{E_A}{kT}\right\}. \qquad \ldots\ldots(3)$$

This equation applies when the electrons in the acceptor levels do not have paired spins (cf. Landsberg 1952). $N_A - N_D$ is equal to the saturation value of the density of holes (2×10^{16} per cm^3). The graph of $\log PT^{-3/2}$ against $1/T$ or $\log RT^{3/2}$ against $1/T$ (figure 4) is now a straight line at low temperatures and its slope gives the activation energy (not one-half of the activation energy as in the simpler case). The result is $E_A = 0{\cdot}38$ ev.

In equation (3) both N_D and m^* are unknown. Using the experimental results it is found that

$$\frac{N_D}{(m^*/m)^{3/2}} = 2 \times 10^{15}\,\text{cm}^{-3}.$$

Therefore, if the effective mass of the holes is equal to the free electron mass, then in this diamond one tenth of the acceptors are filled at absolute zero by electrons from donor levels.

Although several energy levels have been observed between the valence and conduction bands in other types of diamonds, these diamonds are insulators. It is believed that the electrons from the donor levels fill all the lowest acceptor levels, and the only acceptors which are left empty or donors which remain full of electrons are those which lie near the middle of the 6 ev forbidden gap. In Type IIb diamonds the acceptor level at 0·38 ev is not entirely full at room temperature. It appears that there are too few donors to contribute their electrons to this acceptor level. Since Type II diamonds are thought to be purer than Type I diamonds it is unlikely that the conductivity of Type IIb is due to a larger than usual concentration of acceptor levels. (In the measured diamond their density is only 2×10^{16} per cm^3.) It is more probable that the number of donor centres is unusually small and that Type IIb diamonds are even purer than other Type II diamonds.

The hydrogen-like model for a singly ionized impurity atom in diamond (dielectric constant 5·7) leads to an ionization energy of about 0·4 ev, which is close to the measured activation energy of the acceptors in this specimen. This suggests that these centres may be atoms of some dissolved impurity such as trivalent aluminium, which is commonly found in diamonds (Chesley 1942). Vacancies or other crystal imperfections might also act as acceptors in diamond as they do in germanium and silicon (James and Lark-Horovitz 1951) but it is probable that the associated energy levels would be much deeper than the measured activation energy (cf. measurements on irradiated diamonds by Ditchburn *et al.* 1955).

Several other workers have recently measured the electrical properties of Type IIb diamonds. Leivo and Smoluchowski (1955) measured the resistivity of one specimen between room temperature and 425°c and found that the slope of the $(k \log \rho,\ 1/T)$ curve corresponded to an energy of about 0·35 ev. Brophy (1955) measured resistivity and Hall coefficient between −40°c and 60°c. His specimen was also p-type and from the slope of the Hall constant against temperature curve he deduced an activation energy of 0·35 ev. Custers (1955) measured resistance between silver contacts on another diamond at temperatures above 20°c. Using an equation similar to equation (2), he deduced an activation energy of 0·70 ev which again corresponds to a slope of 0·35 ev on the usual $(k \log \rho,\ 1/T)$ graph. Each of these energies was obtained from measurements made near room temperature. In the present experiments, the slopes of the curves in figures 1 and 2 taken near room temperature also lead to energies of 0·35 or 0·36 ev. However, these curves are straight lines only below −50°c and in this region the activation energy is found to be 0·38 ev.

Carrier mobilities have been measured by several methods in non-conducting diamonds (Klick and Maurer 1949, McKay 1950, Redfield 1954, Allemand and Rossel 1954). These methods involve injection of carriers by irradiation with light or charged particles and the results are complicated by the presence of space charge or carriers of both signs. Allemand and Rossel estimated from their crystal counter measurements that the mobility of holes is about $1000\ \mathrm{cm^2\,volt^{-1}\,sec^{-1}}$, or higher, and Redfield's Hall effect measurements led to the result $\mu_H \geqslant 1200\ \mathrm{cm^2\,volt^{-1}\,sec^{-1}}$. Our value of $1550 \pm 150\ \mathrm{cm^2\,volt^{-1}\,sec^{-1}}$ is consistent with these results.

§ 3. Optical Properties

Diamonds show great variation in their infra-red absorption but the characteristic features of the two classes are well established (Blackwell and Sutherland 1949, Sutherland, Blackwell and Simeral 1954). The present specimen shows all the lattice absorption associated with a Type II diamond. In addition there are several sharp peaks and a continuum which, unlike the lattice absorption, are very sensitive to temperature change. The following section presents some measurements on the effect of temperature variation and a preliminary report of photoconductivity associated with this extra absorption.

3.1. *Measurements*

Transmission measurements were made with a Hilger H.800 double beam spectrometer using a rock salt prism. Low temperature measurements were performed with the diamond mounted in a cryostat fitted with rock salt windows. High temperature runs were made with the specimen attached to an electrically heated brass block, mounted in an evacuated vessel.

(*a*) *Transmission measurements.*

The room temperature transmission is shown in figure 5(*a*). Comparison with published data, based on the study of many diamonds (Blackwell and Sutherland 1949) shows that the peaks α, β, γ and δ are not usually found in Type II diamonds. However, Blackwell and Sutherland noted similar peaks in a few Premier Overblue stones. Blackwell (1949) found that many diamonds show negligible absorption in the region 1–2·5 μ, the transmission loss being due to reflection. The present specimen has an absorption coefficient of 3·4 cm^{-1} in the range 1–2·5 μ and this continuum of extra absorption may extend as far as about 3·5 μ. These unusual features have been summarized for convenience in figure 5(*b*).

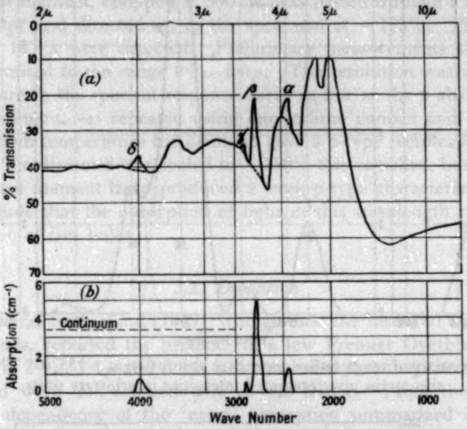


Figure 5. Infra-red absorption in a Type IIb diamond, at room temperature; *a*, absorption spectrum in the region 2–15 μ; *b*, absorption features not usually present in Type II diamonds.

When cooled to -155°c the peaks α, β, γ and δ increase in intensity by about 30%. The main peak β is clearly resolved as a doublet at this temperature and several small peaks appear. Figure 6 shows in detail the extra absorption at low temperature. The lattice absorption has been subtracted and estimates of the relative intensities of the various peaks are given.

When the diamond is heated the α, β, γ and δ peaks decrease in intensity, until at 300°c they are indistinguishable from the background. A slight increase in the transmission level in the region 1–2·5 μ was noted on heating. The temperature variation of the β band was estimated by calculating the integrated absorption (i.e. $\int K\,d\nu$ where K is the absorption coefficient in cm^{-1} and ν the frequency expressed in wave numbers) at various temperatures. Figure 7 shows the results, which have not been corrected for the spectrometer slit width. The slit width amounted to approximately one third of the observed width of the absorption band.

All changes with temperature appeared to be reversible. A slight displacement towards larger wavelengths on heating, amounting to about 3 wave numbers per 100 degrees, was detected in the β band. The half-width of this peak increased by roughly 50% on heating through 100°c, and decreased by about 20% on cooling from room temperature to -155°c.

(*b*) *Homogeneity test.*

The incident light was collimated by means of a pinhole 0·5 mm in diameter and directed through various parts of the crystal. The intensities of the α, β, γ and δ bands remained constant.

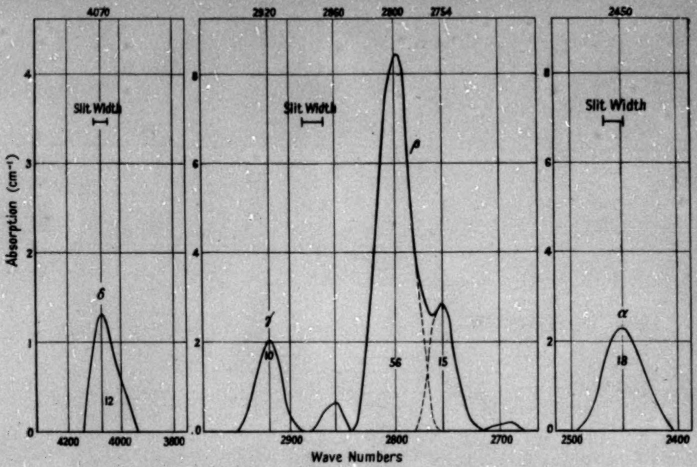


Figure 6. Summary of temperature dependent absorption at -155°C. The figures under the curves give integrated intensities in arbitrary units.

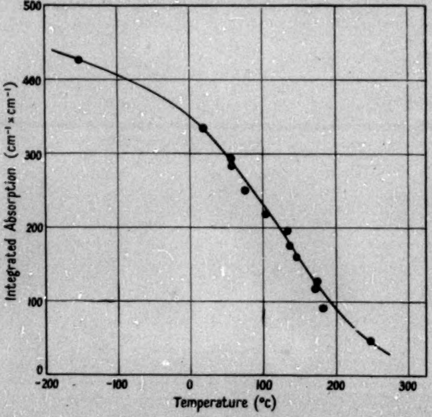


Figure 7. Temperature variation of the integrated absorption of the $3{\cdot}57\,\mu$ peak.

A comparison was made between a transmission T_1 through the length of the crystal (3 mm) and a transmission T_2 through the width (1·5 mm), to determine whether the extra absorption was due to a surface or bulk effect: a plot of $\frac{1}{2}\log T_2$ fitted over $\log T_1$, which indicates that all the optical absorption is a bulk effect.

We conclude that the absorption summarized in figure 5(*b*) is a bulk effect, which is uniform throughout the specimen.

(*c*) *Photoconductivity.*

The diamond was mounted inside the cryostat between pressed indium contacts, and 300 volts were applied in series with a 3 MΩ resistor. Radiation from a Nernst filament, chopped at 500 c/s, was passed through a small fluorite monochromator and directed on to the specimen at −155°c. Photo-currents of the order 10^{-9} A were detected. Preliminary measurements show that the response is limited to the range 0·9 μ–3·6 μ. The resolution was insufficient to detect structure in the spectral response curve as fine as the β absorption peak.

The experiment was repeated using one indium contact and one tungsten probe. At room temperature the diamond gave a p-type rectification characteristic. When the diamond was cooled to −150°c this vanished, but illumination with a tungsten filament lamp produced a weak p-type characteristic.

We conclude that the absorption of light in this wavelength range leads to the production of free holes.

3.2. *Discussion*

Blackwell and Sutherland (1949), who studied the infra-red transmission of many diamonds, reported the presence in a few Premier Overblues of similar sharp bands at 2·5 μ, 3·4 μ and 3·6 μ, and concluded that they were due to some impurity and were not characteristic of the lattice. From the pronounced temperature dependence of the 'extra' absorption summarized in figure 5(*b*) we also conclude that this absorption is not due simply to lattice vibrations, since lattice absorption is relatively temperature independent (Collins and Fan 1954). We suspect that the absorption is characteristic of Type IIb diamonds and is associated with the defect or impurity responsible for conduction.

The optical absorption characteristics of defects in the diamond lattice have been studied by Ditchburn *et al.* (1955) in diamonds subjected to bombardment by electrons and neutrons. They show that complex band systems arise, possibly due to interaction between the electronic configuration of the defects and some lattice vibration. Such complications may well apply here, but since the defect has not been identified no detailed discussion is possible.

However, the similarity of the values of the activation energy of the acceptor (0·38 ev) and the energy of the major temperature dependent absorption peak (0·34 ev) suggests that these may be directly related. The hydrogen-like model leads to the picture of a ground state for holes bound to the acceptor centres and a series of excited states for these holes with energy levels just above the valence band. By analogy with the V_1 colour centres in alkali halides we suggest that the absorption may be associated with transitions from the ground state to an excited state (i.e. transitions of electrons from full levels just above the valence band to the empty acceptor levels, cf. Slater 1949).† It is evident that this absorption should decrease at high temperatures as the acceptor level is filled by thermally excited electrons, but a simple calculation based on the assumption of constant transition probability does not give quantitative agreement with the experimental results shown in figure 7.

† *Note added in proof.* Our attention has been drawn to the optical studies of impurities in silicon by E. Burstein, E. E. Bell, J. W. Davisson and M. Lax, 1953, *J. Phys. Chem.*, **57**, 849. They observed absorption peaks and a continuum in the infra-red (cf. figure 5) which they interpreted in terms of the hydrogen-like model

Using this model, the continuous absorption of shorter wavelength infra-red light and the associated photoconductivity may be explained in terms of the ionization of the acceptors or optical excitation of electrons from the valence band up to the acceptor level. A threshold would be expected at 0·38 ev (3·2 μ) and a peak in the photoconductive response at 0·34 ev (3·6 μ) but the resolution of the monochromator in the present experiments was not sufficient to show this structure.

This interpretation, which is based on the results obtained with a single specimen, is only tentative. It is based on the assumption that the observed infra-red spectrum is characteristic of Type IIb diamonds. This assumption should be verified by observations on other diamonds† with comparable conductivities. Also it would be interesting to determine whether the Premier Overblue diamonds, in which Blackwell and Sutherland (1949) found peaks similar to those in figure 5, are Type IIb diamonds. Further understanding of the connection between the electrical and optical properties of semiconducting diamonds might be obtained by repeating the photoconductivity measurements over a range of temperatures with much better spectral resolution.

ACKNOWLEDGMENTS

We are grateful to Dr. J. F. H. Custers of the Diamond Research Laboratory, Johannesburg, and to Dr. F. P. Bowden and Mr. M. Seal of the University of Cambridge for lending the diamond to us. We are pleased to thank Mr. R. D. Groves for valuable experimental assistance.

† *Note added in proof.* Similar absorption spectra have now been observed in three more Type IIb diamonds.

REFERENCES

ALLEMAND, C., and ROSSEL, J., 1954, *Helv. Phys. Acta*, **27**, 519.

BARDEEN, J., and SHOCKLEY, W., 1950, *Phys. Rev.*, **80**, 72.

BLACKWELL, D. E., 1949, *Thesis*, University of Cambridge.

BLACKWELL, D. E., and SUTHERLAND, G. B. B. M., 1949, *J. Chim. Phys.*, **46**, 9.

BROPHY, J. J., 1955, *Phys. Rev.*, **99**, 1336.

CHESLEY, F. G., 1942, *Amer. Min.*, **27**, 20.

COLLINS, R. J., and FAN, H. Y., 1954, *Phys. Rev.*, **93**, 674.

CUSTERS, J. F. H., 1952, *Physica*, **18**, 489; 1954, *Ibid.*, **20**, 183; 1955, *Nature, Lond.*, **176**, 173.

DITCHBURN, R. W., MITCHELL, E. W. J., PAIGE, E. G. S., CUSTERS, J. F. H., DYER, H. B., and CLARK, C. D., 1955, *Report of Conference on Defects in Crystalline Solids* (London : Physical Society).

DRESSELHAUS, G., KIP, A. F., and KITTEL, C., 1955, *Phys. Rev.*, **98**, 368.

DUNLAP, W. C., 1950, *Phys. Rev.*, **79**, 286.

ISENBERG, I., RUSSELL, B. R., and GREENE, R. F., 1948, *Phys. Rev.*, **74**, 1255 (A).

JAMES, H. M., and LARK-HOROVITZ, K., 1951, *Z. Phys. Chem.*, **198**, 107.

KLICK, C. C., and MAURER, R. J., 1949, *Phys. Rev.*, **76**, 179.

LANDSBERG, P. T., 1952, *Proc. Phys. Soc.* A, **65**, 604.

LEIVO, W. J., and SMOLUCHOWSKI, R., 1955, *Bull. Amer. Phys. Soc.*, **30**, No. 2, 9.

MCKAY, K. G., 1950, *Phys. Rev.*, **74**, 1606.

REDFIELD, A. G., 1954, *Phys. Rev.*, **94**, 526.

ROBERTSON, R., FOX, J. J., and MARTIN, A. E., 1934, *Phil. Trans. Roy. Soc.* A, **232**, 463.

SLATER, J. C., 1949, *Phys. Rev.*, **76**, 1592.

SUTHERLAND, G. B. B. M., BLACKWELL, D. E., and SIMERAL, W. G., 1954, *Nature, Lond.*, **174**, 901.

TROTT, N. G., 1953, *Proc. Roy. Soc.* A, **220**, 498.

WILLARDSON, R. K., and DANIELSON, G. C., 1952, *J. Opt. Soc. Amer.*, **42**, 42.

The Variation with Temperature of the Spectral Emissivities of Iron, Nickel and Cobalt

By L. WARD
University College of the Gold Coast, Accra

MS. received 9*th August* 1955 *and in amended form* 31*st October* 1955

Abstract. The spectral emissivities of iron, nickel and cobalt have been measured over a wavelength range $1{\cdot}0\,\mu$ to $3{\cdot}0\,\mu$ and at temperatures between 650° and 1350°c. The results show anomalies in the cases of iron and cobalt occurring at temperatures in the neighbourhood of the respective Curie points. Tables of the temperature coefficients of emissivity for the three metals are given.

---

§ 1. Introduction

In a previous paper by Lund and Ward (1952) the results were reported of an investigation into the variation with temperature of the spectral emissivities of iron, nickel and cobalt over a wavelength range $1{\cdot}0\,\mu$ to $2{\cdot}6\,\mu$. The most interesting feature of these results was a significant change in emissivity at the Curie point of cobalt (1150°c).

The apparatus used then was not sufficiently sensitive to enable measurements to be made around the Curie point of iron (770°c) and the present investigation was commenced with this objective in view.

§ 2. Apparatus

The apparatus was an improved modification of that used in the earlier work. Drecq-type black bodies (1914) were made by bending flat blanks of the metal into cylindrical form, 6 in. long and ¼ in. in diameter, leaving a longitudinal gap of $\frac{1}{16}$ in. Heating was by a Variac-controlled current transformer and the radiation was measured by a lead sulphide cell in conjunction with a quartz spectrometer.

Details of the water cooled chamber housing the specimens are shown in figure 1. The chief improvements over the earlier model were (*a*) the use of vertical supporting electrodes mounted on a horizontal base-plate, resulting in a much more satisfactory vacuum joint, and (*b*) additional resilience in the specimen mounting provided by the flat spiral copper spring. The chamber was evacuated by a two-stage rotary pump and filled with hydrogen, a getter being used to eliminate the last traces of oxygen.

A Hilger Barfit spectrometer was used, passing a narrow band of wavelengths of about $0{\cdot}03\,\mu$ at each point.

The radiation was focused on the spectrometer by a concave front silvered mirror of focal length 25 cm, the image being roughly the same size as the object. Immediately in front of the spectrometer the radiation beam was interrupted by a sectored wheel, driven by a synchronous motor, to produce an interruption frequency of 750 per second.

The radiation was measured by a lead sulphide photoconducting cell coupled with a tuned audio frequency amplifier. A B.T.H. type CI cell was used at first, but later a Mullard C15V cell was introduced with a five-fold increase in sensitivity.

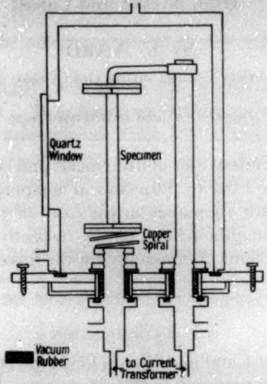


Figure 1. Specimen assembly.

No cooling liquids were available and both cells were operated at room temperature. The signal from a head amplifier was applied to a voltage dividing box from which it was fed to the main amplifier tuned to 750 c/s and was recorded on a micro-ammeter in the anode circuit of the output valve.

The voltage dividing box was used to attenuate the signals to give the same output from the black body and the surface, the ratio of the readings on the box giving the emissivity of the surface. The departure from linearity of response of the detector head unit was studied and it was found that the true emissivity was given by the measured value raised to the power 1·035. In practice this correction could be applied in almost all cases simply by subtracting 0·021 from the measured value.

It was not possible to measure the efficiency of the black body but the greyness was tested by measuring the output θ (reciprocal of voltage dividing box reading) for different temperatures T, and plotting $\ln \theta$ against $1/T$. A straight line at all temperatures indicated that the efficiency of the black body was not changing (see Hall 1944).

When dealing with this type of black body, the condition for maximum efficiency is to obtain a uniform temperature over the central portion equal to at least 8 diameters. An optical pyrometer was used in plotting the temperature distribution and, by thinning down the ends of the specimen with a file, a length to diameter ratio of 10 at constant temperature was easily obtained.

The specimen surfaces were prepared by polishing with emery paper down to grade 04 and finally lapping with jeweller's rouge. The exact crystalline state of the surface was not known, but the amount of cold work done in polishing was relatively small and all the specimens were subjected to the same treatment.

In taking measurements, six values each of the black body and surface readings were made and the mean taken. The departure from the mean rarely exceeded

1% and the emissivity should be accurate to better than ½%. Three specimens of each material were used and the results given are the average of these. The absolute values of emissivity differed by up to 8% amongst the individual specimens.

The method of taking readings was to keep the wavelength fixed and vary the temperature, starting from the highest down to the lowest at which significant values could be taken and back again. In the middle of the wavelength range, readings could be extended down to 650°c but at each end it was difficult to proceed below 750°c.

§ 3. RESULTS

3.1. *Nickel*

The mean results for nickel at various wavelengths are shown in figure 2. Compared with other published work the values are a little high, almost certainly due to the surface condition. The variation with temperature was treated as linear throughout and the temperature coefficients of emissivity referred to 1000°c were evaluated from the graphs and are shown in table 1.

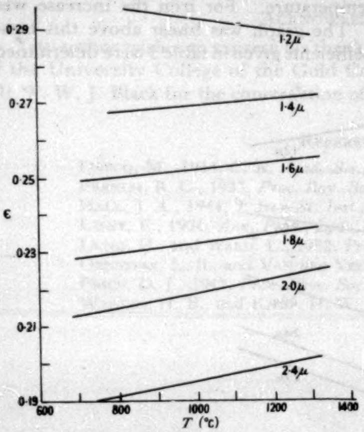


Figure 2. Emissivity of nickel against temperature.

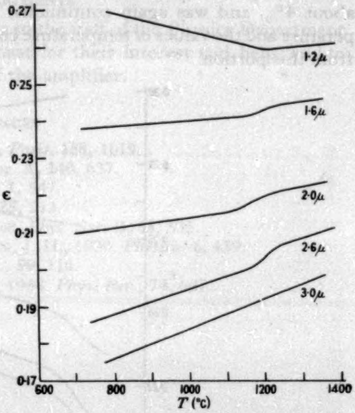


Figure 3. Emissivity of cobalt against temperature.

Table 1

$\lambda(\mu)$	1·0	1·2	1·4	1·6	1·8	2·0	2·2	2·4	2·6	2·8	3·0	3·2
$\alpha \times 10^5$ deg^{-1}	−10	−5	−3	5	7	8	9	11	13	18	19	21

The results indicate an 'X-point' at 1·5 μ, therein differing from the values of Lund and Ward (1952) and Price (1947). A gradual increase in emissivity was noticed with prolonged heating and recrystallization of the surface was observed simultaneously.

3.2. *Cobalt*

The mean values for the three cobalt specimens are displayed in figure 3. It will be seen that, for wavelengths up to 2·6 μ, these graphs show an upwards inflection of about 2% between 1100°c and 1200°c. This range includes the Curie temperature, but it was noticed that the change in emissivity took place gradually over a range of about 50°c and no actual discontinuity could be observed. The curves resumed their original slope above the inflection. The values of the temperature coefficients determined from the linear portion below 1100°c are given in table 2.

Table 2

$\lambda(\mu)$	1·1	1·2	1·4	1·6	1·8	2·0	2·2	2·4	2·6	2·8	3·0
$\alpha \times 10^5$ deg^{-1}	−8	−9	−5	3	6	8	12	15	20	22	23

An 'X-point' is indicated at 1·5μ in good agreement with 1·4μ obtained previously.

3.3. *Iron*

The curves for iron are shown in figure 4. As with cobalt there is an upwards inflection in the region of the Curie temperature. For iron the increase was about 4%, and was again continuous. The graph was linear above this temperature and the values of temperature coefficients given in table 3 were determined from this portion.

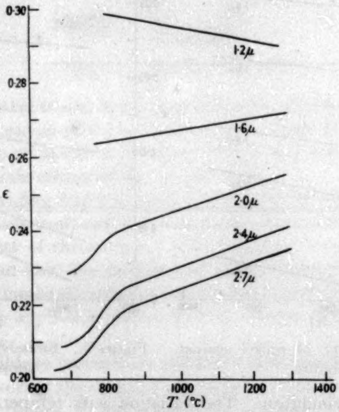


Figure 4. Emissivity of iron against temperature.

Table 3

$\lambda(\mu)$	1·0	1·2	1·4	1·6	1·8	2·0	2·2	2·4	2·7	2·9
$\alpha \times 10^5$ deg^{-1}	−10	−7	−6	6	11	12	14	17	18	20

§ 4. Discussion

In the results presented for iron and cobalt one fact seems to be significant, namely that a fairly abrupt change in emissivity occurs at temperatures near the Curie points. Lowe (1936) reports a similar phenomenon in the case of nickel. The physical cause of these changes is not known as the crystal structure does not change at the Curie temperatures. On the other hand these results do not indicate any significant changes in emissivity at the phase transformation points. Wahlin and Knop (1948) working with iron at 0·65 μ observed a discontinuous decrease in emissivity at 900°c and 1400°c, the phase transformation points, and Ornstein and Van der Veen (1939) found the change at 900°c and also a continuous increase in reflectivity at the Curie point. With cobalt, Wahlin and Knop found a sudden change on passing through the Curie point.

The electrical resistivities of these elements plotted against temperature show inflections at the Curie points, and are related to the emissivities by the Hagen–Rubens equation. The latter breaks down below 4 μ, but some correlation with the resistance presumably remains.

It is perhaps surprising that the crystal structure has no apparent effect on the emissivity at these wavelengths. This may be due to the presence of an amorphous surface layer, although the amount of cold work performed in polishing would hardly have been sufficient to produce a complete breakdown in crystal structure (see French 1933).

ACKNOWLEDGMENTS

The author wishes to express his thanks to the staff of the Physics Department of the University College of the Gold Coast for their interest and help, and to Mr. P. W. J. Black for the construction of the amplifier.

REFERENCES

DRECQ, M., 1914, *C. R. Acad. Sci., Paris*, **158,** 1019.
FRENCH, R. C., 1933, *Proc. Roy. Soc.* A, **140,** 637.
HALL, J. A., 1944, *J. Iron St. Inst.*, **1,** 547.
LOWE, E., 1936, *Ann. Phys., Lpz.*, **25,** 213.
LUND, H., and WARD, L., 1952, *Proc. Phys. Soc.* B, **65,** 535.
ORNSTEIN, L. S., and VAN DER VEEN, J. H., 1939, *Physica*, **6,** 439.
PRICE, D. J., 1947, *Proc. Phys. Soc.*, **59,** 118.
WAHLIN, H. B., and KNOP, H. W., 1948, *Phys. Rev.*, **74,** 687.

Birefringence Resulting from the Application of an Intense Beam of Light to an Isotropic Medium

By A. D. BUCKINGHAM †
Department of Theoretical Chemistry, University of Cambridge

MS. received 19*th October* 1955

Abstract. The interaction between a rapidly oscillating electric field, such as is found associated with a light wave, and anisotropic molecules is discussed, and it is shown that there will be a tendency for the molecules to be oriented. An isotropic medium will therefore become doubly refracting, the optic axis being parallel to the electric vector of the oscillating field, or in the case of normal light to its direction of propagation. For the birefringence to be easily measurable a very intense beam of light would be necessary, and a flash apparatus would probably be the most suitable source.

---

§ 1. INTRODUCTION

KERR (1875) showed that the application of a strong static electric field to an isotropic medium caused it to become doubly refracting, or birefringent. The effect has since been observed in a wide range of substances, and is now known as the Kerr electro-optical effect. The theory of the effect has been discussed by Langevin (1910), Born (1918), Debye (1925) and recently by Buckingham and Pople (1955). The birefringence is due chiefly to the tendency of the external electric field to orient the molecules in such a way that their energy is a minimum. If the molecules are anisotropically polarizable, then there is a tendency for the axis of maximum polarizability to be parallel to the applied field, and the corresponding torque may either add to, or subtract from, the turning moment due to the permanent dipole moments of the molecules; the sign depends on the direction of the permanent moment relative to the principal axes of polarizability. There is also a small temperature independent effect found even in the spherically symmetric inert gases (see Kuss and Stuart 1941), and this is due to the anisotropy produced in the molecule by the external field.

If a parallel beam of light is passed through an assembly of anisotropically polarizable molecules, there will be a tendency for the molecules to be lined up owing to their interaction with the electric field associated with the waves. We suppose that the frequency of the light is sufficiently high for the molecules to be unable to follow the alternations, and such that there is no appreciable absorption. After averaging with respect to time, there will be a resultant interaction energy proportional to the square of the amplitude of the waves of the incident light. However, there will be no term resulting from the interactions of the permanent dipoles with the field, for this vanishes on averaging with respect to the time. The non-zero interaction due to the anisotropy in the polarizabilities of the molecules will tend to orient them in such a way that their

† Now at Department of Physical Chemistry, University of Oxford.

axes of maximum polarizability are at right angles to the direction of propagation of the beam, and the substance will then behave as a doubly refracting medium with optic axis in the direction of propagation. The magnetic field accompanying the light rays will have an insignificant orienting effect on the molecules and can safely be ignored in a theoretical treatment of the phenomenon.

§ 2. THE BASIC EQUATIONS FOR A PERFECT GAS

In measuring the double refraction of a specimen, we effectively determine the difference between the refractive indices $n_{\parallel}$ and $n_{\perp}$ of the sample in the direction of, and at right angles to, the optic axis. In this section, we shall derive the fundamental equations for the double refraction of an assembly of like independent molecules in a rapidly oscillating electric field $\mathscr{E}=\mathscr{E}_0\cos(2\pi\nu t)$, the direction of the electric vector being that of the unit vector $\mathbf{e}^{\perp}$, where $\mathbf{e}^{\perp}$ is a fixed direction at right angles to the direction of propagation of the light beam which we shall represent by the unit vector $\mathbf{e}^{\parallel}$. One half of the waves comprising the beam may be supposed to have electric vectors vibrating in the direction of $\mathbf{e}^{\perp}$, those of the other half being at right angles both to $\mathbf{e}^{\perp}$ and to $\mathbf{e}^{\parallel}$.

In the notation of Buckingham and Pople (1955), we represent the energy of a molecule in the field by $u(\tau, \mathscr{E})$, where τ describes the position and orientation of the molecule at a particular instant. We shall write

$$\begin{aligned} u(\tau,\mathscr{E}) &= u^{(0)}-\int_0^{\mathscr{E}}\mu_\alpha e_\alpha{}^{\perp}\,d\mathscr{E} \\ &= u^{(0)}-\mu_{0\alpha}e_\alpha{}^{\perp}\mathscr{E}-\tfrac{1}{2}\alpha_{\alpha\beta}e_\alpha{}^{\perp}e_\beta{}^{\perp}\mathscr{E}^2-\tfrac{1}{6}\beta_{\alpha\beta\gamma}e_\alpha{}^{\perp}e_\beta{}^{\perp}e_\gamma{}^{\perp}\mathscr{E}^3 \\ &\quad -\tfrac{1}{24}\gamma_{\alpha\beta\gamma\delta}e_\alpha{}^{\perp}e_\beta{}^{\perp}e_\gamma{}^{\perp}e_\delta{}^{\perp}\mathscr{E}^4-\ldots \qquad \ldots\ldots(2.1)\end{aligned}$$

where μ_α is the total dipole moment of the molecule and $\mu_{0\alpha}$, $\alpha_{\alpha\beta}$, $\beta_{\alpha\beta\gamma}$, $\gamma_{\alpha\beta\gamma\delta}$, ... are tensors symmetric in all suffixes. The quantities $\mu_{0\alpha}$ and $\alpha_{\alpha\beta}$ are simply the permanent dipole moment and polarizability tensor of the molecule, while $\beta_{\alpha\beta\gamma}$, $\gamma_{\alpha\beta\gamma\delta}$, ... describe the departure from linear dependence of the induced dipole moment on the field strength. If we suppose that the period of oscillation ν^{-1} is too small for the molecules to follow the field, then the 'effective' energy is found by averaging $u(\tau, \mathscr{E})$ over one complete oscillation. Writing this average with respect to time for a given configuration τ as $[u(\tau, \mathscr{E})]$, we have

$$\begin{aligned} [u(\tau,\mathscr{E})] &= \nu\int_0^{\nu^{-1}} u(\tau,\mathscr{E})\,dt \\ &= u^{(0)}-\tfrac{1}{4}\alpha_{\alpha\beta}e_\alpha{}^{\perp}e_\beta{}^{\perp}\mathscr{E}_0{}^2-\tfrac{1}{64}\gamma_{\alpha\beta\gamma\delta}e_\alpha{}^{\perp}e_\beta{}^{\perp}e_\gamma{}^{\perp}e_\delta{}^{\perp}\mathscr{E}_0{}^4-\ldots. \qquad \ldots\ldots(2.2)\end{aligned}$$

It is this time-averaged energy $[u(\tau, \mathscr{E})]$ that must be inserted into the appropriate Boltzmann type weighting factor when we perform the averaging over configuration space.

The quantity we are concerned with is the difference between the mean polarizabilities of a molecule in the directions $\mathbf{e}^{\parallel}$ and $\mathbf{e}^{\perp}$. Again using the notation of Buckingham and Pople (1955) we write for this difference

$$\pi(\tau,\mathscr{E})=\pi_{\alpha\beta}(e_\alpha{}^{\parallel}e_\beta{}^{\parallel}-e_\alpha{}^{\perp}e_\beta{}^{\perp}) \qquad \ldots\ldots(2.3)$$

where $\pi_{\alpha\beta}$ is the 'differential polarizability tensor', defined by

$$\begin{aligned} \pi_{\alpha\beta} &= \frac{\partial\mu_\alpha}{\partial(\mathscr{E}e_\beta{}^{\perp})}=-\frac{\partial^2 u(\tau,\mathscr{E})}{\partial(\mathscr{E}e_\alpha{}^{\perp})\partial(\mathscr{E}e_\beta{}^{\perp})} \\ &= \alpha_{\alpha\beta}+\beta_{\alpha\beta\gamma}e_\gamma{}^{\perp}\mathscr{E}+\tfrac{1}{2}\gamma_{\alpha\beta\gamma\delta}e_\gamma{}^{\perp}e_\delta{}^{\perp}\mathscr{E}^2+\ldots. \qquad \ldots\ldots(2.4)\end{aligned}$$

If there are N/V molecules in unit volume

$$n_{\parallel}-n_{\perp}=\frac{2\pi N}{V}\overline{[\pi(\tau,\mathscr{E})]} \qquad \ldots\ldots(2.5)$$

where

$$[\pi(\tau,\mathscr{E})]=\nu\int_0^{\nu^{-1}}\pi(\tau,\mathscr{E})\,dt$$

$$=(\alpha_{\alpha\beta}+\tfrac{1}{4}\gamma_{\alpha\beta\gamma\delta}e_{\gamma}{}^{\perp}e_{\delta}{}^{\perp}\mathscr{E}_0{}^2+\ldots)(e_{\alpha}{}^{\parallel}e_{\beta}{}^{\parallel}-e_{\alpha}{}^{\perp}e_{\beta}{}^{\perp}), \qquad \ldots\ldots(2.6)$$

and where the bar in (2.5) signifies an average over all configurations; thus

$$\overline{[\pi(\tau,\mathscr{E})]}=\frac{\int[\pi(\tau,\mathscr{E})]\exp\{-[u(\tau,\mathscr{E})]/kT\}\,d\tau}{\int\exp\{-[u(\tau,\mathscr{E})]/kT\}\,d\tau}. \qquad \ldots\ldots(2.7)$$

We shall now expand $\overline{[\pi(\tau,\mathscr{E})]}$ as a power series in $\mathscr{E}_0$, and shall use the notation $\langle X\rangle$ for the average value of a quantity $[X(\tau,\mathscr{E})]$ in the assembly with $\mathscr{E}_0=0$. That is

$$\langle X\rangle=\frac{\int[X(\tau,0)]\exp\{-u^{(0)}/kT\}\,d\tau}{\int\exp\{-u^{(0)}/kT\}\,d\tau}. \qquad \ldots\ldots(2.8)$$

The leading non-vanishing term in the expansion is the one in $\mathscr{E}_0{}^2$, the coefficient being

$$\frac{1}{2}\left(\frac{\partial^2\overline{[\pi(\tau,\mathscr{E})]}}{\partial\mathscr{E}_0{}^2}\right)_{\mathscr{E}_0=0}=\frac{1}{2}\left\langle\frac{\partial^2[\pi(\tau,\mathscr{E})]}{\partial\mathscr{E}_0{}^2}\right\rangle-\frac{1}{2kT}\left\langle[\pi(\tau,\mathscr{E})]\frac{\partial^2[u(\tau,\mathscr{E})]}{\partial\mathscr{E}_0{}^2}\right\rangle. \qquad \ldots(2.9)$$

It will now be convenient to introduce a function L characteristic of the substance comprising the sample. For any specimen in which the electric intensity is independent of position

$$L=\frac{12nV_{\mathrm{m}}}{(n^2+2)^2\{(n')^2+2\}^2}\lim_{\mathscr{E}_0{}'\to 0}\left\{\frac{n_{\perp}-n_{\parallel}}{(\mathscr{E}_0{}')^2}\right\}, \qquad \ldots\ldots(2.10)$$

where V_{m} is the molar volume, n the refractive index and $\mathscr{E}_0{}'$ the amplitude of the orienting beam's electric vector in the medium. n' is the refractive index of the medium for waves of frequency ν. If $n_{\perp}-n_{\parallel}$ is measured with light of a different frequency, then n and n' may differ on account of the effects of atomic polarization. $\mathscr{E}_0{}'=\mathscr{E}_0\{3/[(n')^2+2]\}$ in the case of a spherical specimen and (2.10) reduces to

$$L=\frac{4nV_{\mathrm{m}}}{3(n^2+2)^2}\lim_{\mathscr{E}_0\to 0}\left\{\frac{n_{\perp}-n_{\parallel}}{\mathscr{E}_0{}^2}\right\}. \qquad \ldots\ldots(2.11)$$

For gases at low densities n is unity, whence

$$L=\frac{4V_{\mathrm{m}}}{27}\lim_{\mathscr{E}_0\to 0}\left\{\frac{n_{\perp}-n_{\parallel}}{\mathscr{E}_0{}^2}\right\}=\frac{4\pi\boldsymbol{N}}{27}\left(\frac{\partial^2\overline{[\pi(\tau,\mathscr{E})]}}{\partial\mathscr{E}_0{}^2}\right)_{\mathscr{E}_0=0} \qquad \ldots\ldots(2.12)$$

where $\boldsymbol{N}$ is Avogadro's number.

From (2.2), (2.6), (2.9) and (2.12), we obtain

$$L=\frac{2\pi\boldsymbol{N}}{27}\left\langle\left\{\gamma_{\alpha\beta\gamma\delta}+\frac{1}{kT}\alpha_{\alpha\beta}\alpha_{\gamma\delta}\right\}\{e_{\alpha}{}^{\perp}e_{\beta}{}^{\perp}(e_{\gamma}{}^{\perp}e_{\delta}{}^{\perp}-e_{\gamma}{}^{\parallel}e_{\delta}{}^{\parallel})\}\right\rangle. \qquad \ldots\ldots(2.13)$$

By a simple trigonometrical analysis, it can be shown that

$$\langle e_{\alpha}{}^{\perp}e_{\beta}{}^{\perp}(e_{\gamma}{}^{\perp}e_{\delta}{}^{\perp}-e_{\gamma}{}^{\parallel}e_{\delta}{}^{\parallel})\rangle=\tfrac{1}{30}\{-2\delta_{\alpha\beta}\delta_{\gamma\delta}+3\delta_{\alpha\gamma}\delta_{\beta\delta}+3\delta_{\alpha\delta}\delta_{\beta\gamma}\}, \qquad \ldots\ldots(2.14)$$

so that

$$L=\frac{2\pi\boldsymbol{N}}{405}\left\{2\gamma_{\alpha\alpha\beta\beta}+\frac{3}{kT}(\alpha_{\alpha\beta}\alpha_{\alpha\beta}-3\alpha^2)\right\}, \qquad \ldots\ldots(2.15)$$

where $\alpha=\frac{1}{3}\alpha_{\alpha\alpha}$ is the mean low-field strength polarizability.

By comparing (2.15) with the expression for the molecular Kerr constant ${}_{\mathrm{m}}K$ deduced by Buckingham and Pople (1955), we see that

$$L = {}_{\mathrm{m}}K - \frac{2\pi \boldsymbol{N}}{405kT}\left\{4\mu_{0\alpha}\beta_{\alpha\beta\beta} + \frac{3}{kT}(\alpha_{\alpha\beta}\mu_{0\alpha}\mu_{0\beta} - \alpha\mu_0{}^2)\right\} \qquad \ldots\ldots(2.16)$$

where we have neglected the effect of atomic polarizability on ${}_{\mathrm{m}}K$. Hence, for a gas consisting of molecules without permanent dipole moments (i.e. $\mu_{0\alpha} = 0$), $L = {}_{\mathrm{m}}K$.

For molecules with spherical symmetry (2.15) becomes

$$L = 4\pi \boldsymbol{N}\gamma/81, \qquad \ldots\ldots(2.17)$$

and γ is such that the moment of a molecule in a strong field E is $(\alpha E + \frac{1}{6}\gamma E^3 + \ldots)$ in the direction of E.

§ 3. The Effect in a Dense Medium

The results of § 2 will only be strictly applicable to gases at low densities, for the molecules were assumed to be independent. In this section, the theory will be extended to dense media, but since the problem is analogous to that of the Kerr effect previously discussed by Buckingham (1955), the results will be presented tersely.

We shall only consider specimens in which the light beam is parallel and uniform. For a sample consisting of molecules which are spherically symmetric when isolated, we find that

$$L = {}_{\mathrm{m}}K = 4\pi \boldsymbol{N}\gamma/81. \qquad \ldots\ldots(3.1)$$

Equation (3.1) applies at all densities, and depends for its validity upon the fact that the mean local field acting on a molecule is $\mathscr{E}\{1 + \mathrm{O}(\alpha^2)\}$, and since α is usually small, this field is effectively equal to $\mathscr{E}$.

A discussion similar to that employed in the treatment of the Kerr effect leads to the result

$$L = \frac{2\pi \boldsymbol{N}}{405kT}\sum_{i=1}^{N}\langle 3\alpha_{\alpha\beta}{}^{(1)}\alpha_{\alpha\beta}{}^{(i)} - \alpha_{\alpha\alpha}{}^{(1)}\alpha_{\beta\beta}{}^{(i)}\rangle, \qquad \ldots\ldots(3.2)$$

for an assembly of molecules whose polarizabilities are independent of the field strength; $\alpha_{\alpha\beta}{}^{(i)}$ is the polarizability tensor of the ith molecule. Equation (3.2) is accurate to the order of α^2, but it omits terms proportional to α^3. On writing (3.2) in terms of the three principal components of polarizability α_1, α_2 and α_3 it becomes

$$L = \frac{2\pi \boldsymbol{N}}{135kT}\left\{\sum_{s=1}^{3}(\alpha_s)^2 - 3\alpha^2 + \sum_{s,t=1}^{3}\alpha_s\alpha_t\iint(\cos^2\theta_{st} - \tfrac{1}{3})n_2(\mathbf{r},\omega)\,d\mathbf{r}\,d\omega\right\} \qquad \ldots(3.3)$$

where θ_{st} is the angle between the s-principal axis of molecule 1 and the t-principal axis of another molecule whose position and orientation are represented by $(\mathbf{r},\omega)$ referred to molecule 1 fixed at the origin, and where $n_2(\mathbf{r},\omega)$ is the probability of there being a molecule in the volume element $d\mathbf{r}$ and orientational element $d\omega$. For axially symmetric molecules, (3.3) reduces to

$$L = \frac{2\pi \boldsymbol{N}\kappa^2\alpha^2}{45kT}\left\{2 + 3\iint(\cos^2\theta - \tfrac{1}{3})n_2(\mathbf{r},\omega)\,d\mathbf{r}\,d\omega\right\} \qquad \ldots\ldots(3.4)$$

where $\kappa = (\alpha_3 - \alpha_1)/3\alpha$ (the axis of the molecule being in the 3 direction) is a measure of the anisotropy of the molecule, and where θ is the angle between the axes of the molecules.

We note that $L = {}_{\mathrm{m}}K$ at all densities (to the order of accuracy of our results, which neglect higher powers of α) for molecules with centres of inversion, but for molecules with permanent dipole moments the two quantities will differ.

From (3.3) we see that for molecules whose interaction energies are independent of orientation, L is independent of density and given by

$$L = \frac{2\pi \boldsymbol{N}}{135kT}\{\alpha_1{}^2 + \alpha_2{}^2 + \alpha_3{}^2 - 3\alpha^2\}. \qquad \ldots\ldots(3.5)$$

§ 4. DISCUSSION

The results of §§2 and 3 indicate that measurements of the double refraction resulting from the application of an intense beam of parallel light will be a valuable supplement to Kerr constant and light scattering data for information concerning the polarizabilities of molecules. Thus, from (2.16), if ${}_{\mathrm{m}}K - L$, $\alpha_{\alpha\beta}$ and $\mu_{0\alpha}$ are known, an estimate of the magnitude of the tensor $\beta_{\alpha\beta\gamma}$ could be made.

In an actual experiment whose object is to investigate this new effect, the measuring techniques would be similar to those employed in a measurement of the Kerr constant. The chief difficulty is likely to be in the production and maintenance of the intense source of parallel light. The amplitude of the electric vector of a light beam of energy density 1 watt cm^{-2} is only 27·5 volt cm^{-1}, and since $\mathscr{E}_0$ is proportional to the square root of the energy density, one would need a beam of the order of 10^5 watts cm^{-2} in order to obtain a readily measurable phase difference, if one uses a cell similar to those employed in observations of the Kerr effect together with existing detection techniques. Such intensities are normally available only as flashes. However, there is no limit to the length of the specimen, and the linearly polarized beam used for determining the birefringence could in theory be passed through the cell any number of times. Thus in this way it may be possible to make measurements utilizing easily obtainable light sources whose energy output is constant.

The optic axis of the specimen through which the intense light beam is passed is in the direction of propagation, so that the weak linearly polarized beam for measuring the double refraction may have any direction at right angles to this axis, but its plane of polarization should be at an angle of 45° to it. Any experimental difficulties which might arise as a result of light being scattered by the molecules could be avoided by the use of light of different frequencies for the measuring and orienting beams, together with appropriate filters. In this case, the formulae of §§2 and 3 would require slight modification to allow for possible differences in the polarizabilities of molecules for the two frequencies. Thus (3.3) would be changed to

$$L = \frac{2\pi \boldsymbol{N}}{135kT}\left\{\sum_{s=1}^{3} \alpha_s a_s - 3\alpha a + \sum_{s,t=1}^{3} \alpha_s a_t \int\int (\cos^2\theta_{st} - \tfrac{1}{3}) n_2(\mathbf{r}, \omega)\, d\mathbf{r}\, d\omega\right\} \qquad \ldots\ldots(4.1)$$

where α_s and a_s are the principal polarizabilities for the two frequencies.

ACKNOWLEDGMENTS

The author is indebted to Dr. J. A. Pople and Dr. B. J. Alder for some helpful suggestions concerning the foregoing topics, and to the Royal Dutch/Shell Group for the award of a Postgraduate Scholarship.

REFERENCES

BORN, M., 1918, *Ann. Phys., Lpz.*, **55**, 177.
BUCKINGHAM, A. D., 1955, *Proc. Phys. Soc.* A, **68**, 910.
BUCKINGHAM, A. D., and POPLE, J. A., 1955, *Proc. Phys. Soc.* A, **68**, 905.
DEBYE, P., 1925, *Handbuch der Radiologie*, **6**, 760.
KERR, J., 1875, *Phil. Mag.*, **50**, 337, 446.
KUSS, E., and STUART, H. A., 1941, *Phys. Z.*, **42**, 95.
LANGEVIN, P., 1910, *Radium, Paris*, **7**, 249.

The Scattering of Light by Single Crystals

By S. P. F. HUMPHREYS-OWEN
Birkbeck College, University of London

MS. received 1*st November* 1955, *and in final form* 12*th December* 1955

Abstract. The intensity of light scattered at 90° to the incident beam has been investigated as a function of wavelength for a specimen of colourless, apparently perfect, quartz. The wavelength function indicates that, in addition to the light scattered by thermal fluctuations of density, to be expected in any crystal, however perfect, there is a contribution from some real-structure possessing a linear dimension of about 3000 Å. Previous work by the present author pointed to a similar effect in NaCl crystals, in which a dimension of about 1500 Å was derived, and a preliminary interpretation of the effect was offered in terms of a theoretical real-structure first suggested by Born and later discussed by Fürth. It is now pointed out that the formula given by Born does lead to a dimension for quartz about twice that for NaCl. But since more than one kind of imperfection could produce this 'size effect' a discussion is given of imperfections from this point of view. Two kinds come into consideration: (i) clusters of the correct size randomly embedded in the lattice, (ii) a three-dimensional net throughout the lattice, made up of surfaces of disorder enclosing regions, of the correct size, of perfection. It is argued that the former will produce discrete appearances in the beam of light, detectable by ultra-microscope, and that an entirely diffuse beam of light indicates their absence. This is the case with several crystals which nevertheless exhibit the size effect, and it seems therefore that the latter alternative, the net, has to be accepted. This conclusion is discussed with reference to present knowledge of subgrain structure as well as to the theory of Born and Fürth.

---

§ 1. INTRODUCTION

THE author, in a previous paper (1955), investigated the intensity of light scattered transversely by NaCl single crystals as a function of wavelength, and found that the results were not in accordance with the theory for perfect material. This theory requires that the scattered intensity be due entirely to thermal fluctuations of density, of dimension small compared with the wavelength, and that therefore it be proportional to the inverse fourth power of the wavelength (λ^{-4} law). The observed wavelength function was different, and indicated that there were scattering centres in the crystal of size comparable with the wavelength.

A model was proposed in which the scattering was said to be due to the summation of wavelets scattered from disordered boundaries, the boundaries being the inter-surfaces of regions within which the lattice was assumed to be perfect. The shape of the perfect regions was said to be compact but irregular, and the total effect was assumed to approximate to a random assembly of dielectric spheres of refractive index equal to that of the crystal. There would also be thermal scattering from within the perfect regions, but in the particular cases studied this was shown to be an unimportant fraction of the total scattering.

Computed tables (see references) for the scattering by dielectric spheres of refractive index 1·55 were utilized. From these tables the total luminous flux scattered per unit solid angle by a sphere of radius r, irradiated by unpolarized light of wavelength λ and unit intensity, can be obtained for various angles to the incident beam and for various values of the parameter $\alpha = 2\pi r/\lambda$. A feature of 90° scattering by spheres is a maximum of intensity at $\alpha = 1{\cdot}8$. Thus a maximum at a certain wavelength allows a linear dimension, or equivalent diameter, $2r$ to be inferred.

The experimental curves for NaCl agreed rather well with the theoretical curves for spheres, not only in the observation of maxima in certain cases, but also in the general form of the curves, and indicated the presence of a quite sharply defined linear dimension of about 1500 Å. This dimension did not vary much from crystal to crystal, and it was suggested that there was evidence here in favour of a theoretical real-structure first proposed by Born (1947) and later discussed by Fürth (1949). But, as already pointed out by Theimer and Lell (1953), more than one type of imperfection in a crystal may scatter light in a manner deviating from the λ^{-4} law, and it is not correct to cite departure from the λ^{-4} law as conclusive evidence for a particular real-structure. However, the small variation in the derived dimension, and the absence of visible imperfections in the illuminated volume of the crystals, was thought to be evidence for something intrinsic rather than accidental.

In the present paper a marked departure from the λ^{-4} law is reported also for a specimen of good colourless transparent quartz. But before adducing this as new evidence in favour of the suggested real-structure, it has been thought advisable to review various kinds of slight imperfection and to consider whether any of them could serve as explanation for the observed effects. This review, given in § 4 below, mentions localized clusters of imperfection which would scatter light relatively intensely, and which might have sizes comparable with the wavelengths employed, and thus produce the observed intensity/wavelength effects. But they would also be detectable visually by ultra-microscope, in whose field they would appear as discrete points of light.

Therefore the interior of all crystals has been examined by microscope transversely to a beam of intense light passed through the crystal. Only those crystals in which no trace of localized appearances can be seen have been taken into consideration. With them, illumination of the field of vision, though not always uniform, is entirely diffuse.

§ 2. EXPERIMENTAL

The apparatus and procedure is basically that already described in the author's 1955 paper, but certain changes have been necessary, caused by the lower intensity of light scattered by the quartz specimen. In the previous work the comparison between scattered light and incident light was made by means of a null method. This gave increased convenience and accuracy but was practicable only because the NaCl specimens scattered sufficiently. The method depended, for sensitivity, on detecting the difference of two nearly equal quantities—photocurrent from scattered light, and controlled photocurrent from a source proportional to incident light. The sensitivity becomes low if the scattered light is feeble, because the difference of two nearly equal *small* quantities has to be detected. This was the case in the present work on quartz, and the comparison between scattered and incident light had to be done in separate experiments. In the event, however, the decrease in accuracy was tolerable.

Secondly, it was found in the previous work that parasitic light from the polished NaCl surfaces was an insignificant fraction of the light scattered from the interior. But with the quartz, in the work now described, the parasitic light was appreciable despite the high polish possible of a quartz surface. In view of the intention to carry the observations as far as possible into the ultraviolet, it was decided not to immerse the crystal in a liquid such as benzene of similar refractive index. Instead, end pieces were affixed by 'optical contact' to the quartz block. This expedient, see figure 1, sets back the entrance and exit surfaces so that diffuse reflections from them cannot reach the viewing surface.

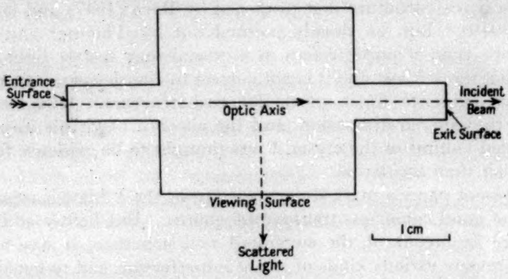


Figure 1. Block of quartz with affixed end pieces in order to prevent diffusely reflected light from entrance and exit surfaces from reaching the viewing surface. All surfaces except these three are painted black.

§ 3. RESULTS

3.1. *Ultra-Microscopic Examination of Crystals*

All the NaCl crystals of the previous paper, and the quartz crystal of this paper, have been examined ultra-microscopically in order to detect localized, light-scattering, imperfections. Only one of them has been found to contain them; this is the synthetic solution-grown NaCl crystal designated in that paper as crystal No. 4. This crystal scattered light 20 times more intensely than the others, and also showed a departure from λ^{-4} law. This bears out a conclusion drawn later in the present paper, in § 4, that *localized* imperfections which are large enough to depart from λ^{-4} law are also relatively intense scatterers. Further evidence on this point has been obtained in the examination of synthetic melt-grown crystals (method of Bridgman) of KCl, KBr and LiF. (It had been hoped to extend the results obtained with the NaCl crystals to other alkali halide crystals, but so far none have been found which are free from ultra-microscopically visible imperfections.)

The scattering of these crystals is relatively intense, of the same order as that of the solution-grown NaCl crystal mentioned above. Ultra-microscopic examination reveals a dense system of points of light, arranged in an interesting manner in the neighbourhood of subgrain boundaries and other complex dislocation systems. It is hoped to publish photographs and further discussion on another occasion. These points of light almost certainly arise from clusters of impurity atoms. The manufacturer has informed the author that impurity, much of which collects into an opaque boule in the growing crucible, has been very troublesome.

Intensity/wavelength study of these crystals shows departure from λ^{-4} law, but fluorescence in the near ultra-violet prevented an estimate of equivalent diameter. The clusters, nevertheless, must be of an average size comparable with the wavelength. But the point to be made is that their visual appearance under the ultra-microscope and their intense scattering power prevents them from being confused with the structure actually under investigation. Although both structures show departure from λ^{-4} law, the latter is distinct in its relatively feeble scattering power and its entirely diffuse appearance ultra-microscopically.

3.2. *Intensity/Wavelength Results for the Quartz Crystal*

Curve A of figure 2 shows a quantity proportional to scattered flux per unit incident intensity plotted against λ^{-4}. There is clearly departure from the λ^{-4} law, under which the curve should be a straight line through the origin. But there is, nevertheless, an appreciable contribution from thermal scattering, and in fact curve A represents a superposition of real-structure scattering on thermal scattering. The scattered flux was compared, at a wavelength of 6000 Å, with that from a sample of benzene contained in a vessel of the same cross-sectional dimensions as the main quartz block. There was a slight difference between the illuminated volumes viewed, since the end pieces attached to the quartz block

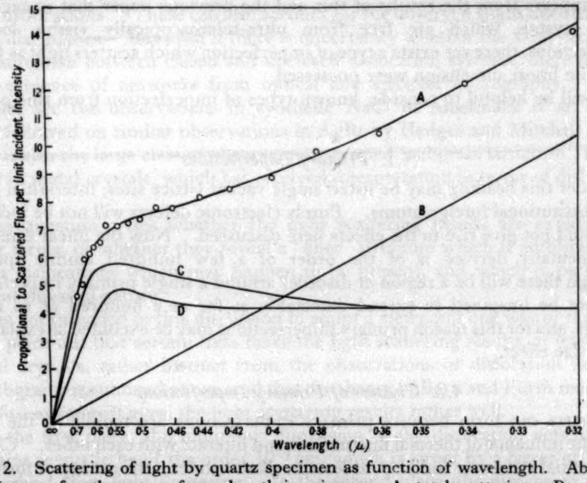


Figure 2. Scattering of light by quartz specimen as function of wavelength. Abscissa is inverse fourth power of wavelength in microns. A, total scattering ; B, estimated thermal scattering; C, real-structure scattering; D, theoretical curve for spheres of diameter 3000 Å, with ordinate scale adjusted for coincidence with curve C at short wavelengths.

modified the optical geometry. But it is not considered that this is large enough to affect the conclusions drawn. The flux ratio, quartz to benzene, was measured as 0·045.

Chapelle and Taurel (1955) give this ratio, based on purely thermal theory, as 0·0065. So at 6000 Å this quartz specimen is scattering about seven times more intensely than if the scattering were entirely thermal.

A straight line representing thermal scattering has therefore been drawn. This is curve B. It passes through the point, at 6000 Å, which has an ordinate value one seventh of that of curve A.

After subtracting the thermal curve B from curve A, curve C is obtained, which represents the real-structure scattering. There is a maximum at 5200 Å, and if this be taken to correspond to $\alpha = 1{\cdot}8$ (see § 1 above) an equivalent diameter of 3000 Å is derived.

A theoretical curve, curve D, for spheres of refractive index 1·55 and diameter 3000 Å is also shown in figure 2. The scale of the ordinates is adjusted so that the level portions beyond 3600 Å coincide. There is satisfactory correspondence in form between the theoretical curve and the experimental curve.

§ 4. DISCUSSION

It appears from the results of this and the previous paper that in regions of single crystals which are free from ultra-microscopically visible localized imperfections, there yet exists a type of imperfection which scatters light as though a definite linear dimension were possessed.

It will be helpful to consider known types of imperfection from this point of view.

4.1. *Primary Imperfections*

Under this heading may be listed single vacant lattice sites, interstitial atoms, and substitutional foreign atoms. Purely electronic defects will not be included; they could not give rise to the effects here discussed. Now the linear dimension experimentally derived is of the order of a few hundred atomic spacings. Although there will be a region of disorder around a single primary imperfection, it cannot be imagined to extend detectably as far as a hundred or so atomic spacings, and for this reason primary imperfections may be excluded as explanation for the size effect.

4.2. *Clusters of Primary Imperfections*

Clusters can form because primary imperfections are mobile in the lattice under the influence of thermal fluctuations and interact with each other. Clusters of interstitial atoms are unlikely, at least in alkali halides; the energy of formation of even a single interstitial atom renders it rare (Seitz 1952, p. 34). Clusters of vacant lattice sites are possible, but these would approximate to voids in the lattice and would produce relatively intense scattering if they were at the same time large enough to produce the size effect. This size also precludes there being so many of them per unit volume that the ultra-microscopic appearance becomes diffuse; such a crystal would be greatly reduced in density. Clusters of foreign atoms are possible and will be visually detectable if there is sufficient optical discontinuity between them and the host lattice. This is certainly the case in the melt-grown alkali halide crystals described in the previous section. But if the cluster formed a lattice of its own, with small optical discontinuity between it and the host lattice, it might evade ultra-microscopic detection. This suggestion is difficult to accept, and in any case it is sufficient for the present argument to emphasize the very uniform dimension derived in the light scattering experiments. There is no reason to expect foreign clusters to be all nearly of the same size.

4.3. *Dislocations*

Lines and surfaces associated with dislocations can act as sinks for primary imperfections, single or in clusters. But the arguments already given, which rule out primary imperfections, or clusters thereof, as explanation of the size effect, apply here.

4.4. *Dislocation Networks*

This is the last remaining possibility to explain the size effect. An early proposal for this kind of structure was that of Darwin (1914) who postulated a 'mosaic' structure in order to account for an apparent breakdown in long-range order in the interpretation of x-ray patterns. This view has lost favour in recent years. For example Guinier (1952) points out that the full picture of a closed mosaic is not essential, and that it is enough to postulate some breakdown of coherence at intervals of about 10^{-4} cm, which can be provided by relatively sparse dislocations. These can also account for the observed slight disorientation of reticular planes. But even if there is uncertainty whether the x-ray technique can distinguish between closed and unclosed dislocation systems, there is much direct evidence of networks from optical and electron micrography. Recent examples are the observations in synthetic NaCl of Amelinckx *et al.* (1955), which followed on similar observations in AgBr by Hedges and Mitchell (1953). There is also the large class of phenomenon, termed 'subgrain structure', studied mostly in metal crystals, which has received interpretation in terms of dislocation theory.

The question arises whether the light scattering results are explicable in similar terms, or whether they reveal a 'finer' structure which is additive to any coarser dislocations which may happen to be present, and which is perhaps of different physical nature.

It is not intended in this discussion to adopt a viewpoint one way or the other, but to point out that certain facts make the light scattering results, at least in the natural crystals, rather distinct from the observations of dislocation networks and subgrains mentioned above, and that the theory of Born and Fürth mentioned in § 1 fits, as far as it goes, the light scattering results rather well.

In the first place, the minimum dimension of the dislocation networks and subgrains seems to be of the order 10^{-4} cm, which is larger by a factor of 10 than that derived from light scattering. In the second place, the dimension derived from light scattering is more uniform than that of the subgrains. In the third place, and perhaps more significantly, the contemporary view is that the subgrains are created only after previous deformation, intentional or accidental, and that they are less likely to appear in natural crystals, particularly if these have not been formed from the liquid. It is noteworthy in this connection that Guinier (1952, p. 418) refers to quartz as a natural crystal which is perfect, it having been formed after a transformation in the solid state.

In the theory of Born and Fürth there is said to be an intrinsic limit to the linear extent of perfection in a crystal, of the order of a few hundred atomic spacings, so that there is spatial periodicity with the order breaking down at intervals with this dimension. The degree of disorder is not specified and therefore the absolute intensity of scattered light can vary without much alteration in the size-dependence of the scattered light. This is in accordance with the experimental findings. The dimension of a few hundred atomic spacings agrees better with the light scattering results than that of the subgrains, but the agreement (see below) is not exact.

The structure is said to be intrinsic. This claim would be invalidated if a perfect crystal were found, but the required accuracy cannot be specified because one cannot define just how near to perfection a crystal with the Born–Fürth structure may be allowed to approach. At least one crystal, a synthetic NaCl specimen, is reported (Taurel and Chapelle 1955) to scatter light in accordance with thermal theory as regards intensity and polarization. This crystal (private communication from Professor Chapelle) does not exhibit this 'perfection' thoughout its volume, there being regions which show excess scattering, but nevertheless volumes of a few mm^3 can be found in which the scattering agrees with thermal theory within the error of the experiments.

The estimate of the limiting dimension l at T°K given by Born is

$$l = \frac{2a}{\beta\theta} \tanh \frac{\theta}{2T},$$

where a is the lattice constant, β the coefficient of linear expansion, and θ the Debye characteristic temperature. It is worth pointing out that this formula does predict a dimension at room temperature for quartz which is about twice that for NaCl, in agreement with the experimental findings. Putting in values for NaCl: $a = 5{\cdot}6 \times 10^{-8}$ cm, $\beta = 4 \times 10^{-5}$ deg^{-1}, $\theta = 281$°K, we get $l = 440$ Å. For quartz, ignoring anisotropy and taking mean values of the relevant quantities, we have $a = 5{\cdot}1 \times 10^{-8}$ cm, $\beta = 1{\cdot}1 \times 10^{-5}$ deg^{-1}. The Debye temperature calculated from elastic constants is 650°K, from specific heat is 730°K, and from reststrahlen is 717°K. Taking 700°K as a mean value we get $l = 1100$ Å, which is 2·5 times that calculated for NaCl. The experimental values were 1500 Å for NaCl and 3000 Å for quartz, so that, although the relative values agree with the theory, the absolute values are three times too high. Nothing can yet be said on this point, since Born's discussion makes no claim beyond order of magnitude.

The present author intends to investigate the temperature dependence of the dimension, but in the meanwhile attention can be drawn to a reversible temperature dependence of scattered intensity, but using white, not monochromatic, light, found by Taurel and Chapelle in the case of another NaCl crystal studied in their 1955 paper. This temperature dependence was quantitatively not in accordance with thermal theory, and is attributed by the authors to some real-structure. This particular observation is also commented on by Fürth (1956), from the point of view of his theory.

ACKNOWLEDGMENTS

Thanks are due to the Council of the Royal Society for a grant to Dr. Fürth from the Scientific Investigations Grant-in-aid, to the Central Research Fund of the University of London for a grant to the author for optical equipment, and to Mlle L. Taurel of the University of Nancy for cooperation in the ultramicroscopic examination of crystals, and for suggesting this procedure in the first place.

REFERENCES

AMELINCKX, S., VAN DER VORST, W., GEVERS, R., and DEKEYSER, W., 1955, *Phil. Mag.*, **46,** 450.
BORN, M., 1947, *Proc. Math. Phys. Soc. Egypt*, **3,** 35.
CHAPELLE, J., and TAUREL, L., 1955, *J. Chim. Phys.*, **52,** 307.
DARWIN, C. G., 1914, *Phil. Mag.*, **40,** 1227.
FÜRTH, R., 1949, *Phil. Mag.*, **27,** 315 ; 1956, *Proc. Phys. Soc.* B, **69,** 255.
GUINIER, A., 1952, *Imperfections in Nearly Perfect Crystals*, 1950, Ed. W. Shockley (New York : J. Wiley), p. 402.
HEDGES, J. M., and MITCHELL, J. W., 1953, *Phil. Mag.*, **44,** 223.
HUMPHREYS-OWEN, S. P. F., 1955, *Proc. Phys. Soc.* B, **68,** 325.
SEITZ, F., 1952, *Imperfections in Nearly Perfect Crystals*, 1950, Ed. W. Shockley (New York : John Wiley), p. 3.
TAUREL, L., and CHAPELLE, J., 1955, *C. R. Acad. Sci., Paris*, **240,** 2507.
THEIMER, O., and LELL, E., 1953, *Canad. J. Phys.*, **31,** 1170.
U.S. NAT. BUR. STAND., 1948, *Appl. Maths. Ser.*, No. 4 (Tables of Scattering Functions for Spherical Particles).

Scintillation Decay Times of Organic Crystals

By G. T. WRIGHT†
Physics Department, Rhodes University, Grahamstown, South Africa

MS. received 2nd June 1955, *and in amended form 9th November* 1955

Abstract. A new experimental technique is described for measuring scintillation decay times in the millimicrosecond region. The method developed uses the fact that a pulsed current, as delivered by the photomultiplier scintillation counter, produces a far greater power dissipation in a resistive load than does the same mean steady current. By measuring the power dissipated the pulse duration, and hence the scintillation decay time, may be found. A particular advantage of this method is that stochastic fluctuations of events in the scintillation counter may be neglected. These fluctuations give the multiplier output pulse a very irregular profile and are analysed in detail using statistical generating functions.

The shapes and durations of scintillations excited in anthracene, *trans*-stilbene and *para*-terphenyl crystals are investigated. For anthracene the scintillation pulses delivered at the multiplier anode are exponential with decay times of 31 mμsec and 53 mμsec for electron and alpha particle excitation respectively. The shapes of these pulses are corrected for the effects of transit time spread in the multiplier and for the effect of self-absorption of fluorescence in the crystal to determine the true shape of the molecular scintillation. For both electron and alpha particle excitation this consists of a short initial spike of fluorescence followed by a long decay, non-exponential, component; the major part of the emission is contained in the long decay component. A qualitative analysis of the measurements for *trans*-stilbene and *para*-terphenyl shows that the scintillations from these crystals are also of this general form.

These results are interpreted in terms of the description proposed previously by the author for the scintillation process in organic crystals. This description predicts this shape for the molecular scintillation and is therefore supported by the present measurements.

§ 1. Introduction

The decay times of fluorescence from organic crystals are extremely short. They are consequently very difficult to measure and it is only recently that reliable techniques have become available for investigating photofluorescence decay times (cf. Bailey and Rollefson 1953) and the decay times of scintillations excited by electrons (cf. Phillips and Swank 1953). For organic systems such measurements are of great importance since they provide much useful information about the fundamental mechanisms of fluorescence. It is generally believed that both fluorescence and scintillation decays are exponential, indicating first order processes, but that the latter are generally much longer than the former; both are increased by any self-absorption of fluorescence in the crystal. The present measurements provide further information about the shape and duration of scintillation pulses excited in anthracene crystals by electrons and alpha particles and in *trans*-stilbene and *para*-terphenyl crystals by alpha particles and support the description of the scintillation process proposed previously by the author (Wright 1955 a).

---

†Now at Mullard Research Laboratories, Salfords, Redhill, Surrey.

A new method for measuring scintillation decay times has been devised for the present investigation. This method is based on the fact that a pulsed current, as delivered by the scintillation counter, produces a far greater power dissipation in the resistive load than does the same mean steady current. By measuring the power developed the duration of the scintillation pulses may be obtained. This method is quick and accurate and possesses two very considerable advantages over previous methods. These are that only very simple apparatus is required and the same apparatus may be used for measuring photofluorescence decay times as well as decay times of scintillations excited by both electrons and alpha particles.

§ 2. THEORETICAL

2.1. *Scintillation Decay Times*

Previous measurements by other experimenters indicate that the scintillation pulse from pure organic crystals is approximately exponential at the multiplier anode. For convenience this discussion will therefore be based on a decay curve of exponential shape.

If transit time spread in the multiplier is negligible each individual ionizing particle absorbed in the crystal produces an exponential pulse at the multiplier anode. This pulse is partially smoothed by the stray capacity of the anode circuit. (The pulse shape is also affected by lead inductances but a detailed calculation for the multiplier used in these measurements, an RCA type 5819, showed that the effects of these are very small.) Thus the effective anode circuit consists of a capacity C in parallel with the resistive load R. This is shown in figure 1 which also illustrates the smoothing action of this circuit. The power developed in the resistance by exponential input pulses of decay time τ, and repetition frequency Z (where $Z \ll 1/\tau$ so that the pulses do not overlap) is:

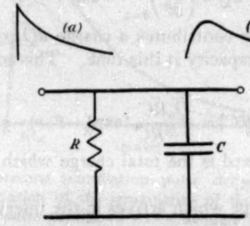


Figure 1. Equivalent anode circuit of the photomultiplier. R, resistive load ; C, stray capacity; (a) shape of input pulse, (b) shape of output pulse.

$$W = i^2 R / 2Z(\tau + RC) \qquad \ldots\ldots(1)$$

and is larger than the steady current value $W = i^2 R$. This is the basic equation used for the measurement of scintillation decay times. Various modifications are required, however, before it may be used in practice: (a) Statistical fluctuations in the rate of emission of photoelectrons from the cathode of the photomultiplier during the scintillation produce an output pulse of irregular shape.

The power developed by this pulse differs from that given by a smooth exponential. (*b*) Due to a variety of causes (Garlick and Wright 1952, Wright and Garlick 1954, Wright 1954 a, b, c) successive scintillation pulses vary in magnitude even when the phosphor crystal is receiving mono-energetic particles.

The effects of these variations are treated most readily by using statistical generating functions to analyse the individual scintillation pulses.

Statistics of scintillation pulses.

We first consider the crystal to be receiving mono-energetic ionizing particles so that successive scintillations have the same mean intensity, and discuss (*a*) above. If the optical collection geometry for the photons of each scintillation is the same then each photon has the same chance p of producing a useful photoelectron, i.e. one which enters the multiplier (Wright 1954 a). If the mean number of photons in each scintillation is $\bar{n}$ the mean number of useful photoelectrons is $\bar{n}p$ and the probability of obtaining np in any given scintillation is $P(np)$ (the Poisson distribution with mean $\bar{n}p$). During the scintillation random fluctuations occur in the rate of emission of photoelectrons. The mean number of these produced in the interval dt at time t after the absorption of the particle is $\bar{m}=\bar{n}p\exp(-t/\tau)\,dt/\tau$. Thus the probability that m useful photoelectrons are produced in this interval is $P(m)$. We may now evaluate the generating function describing the arrival of charge at the multiplier anode during the interval dt. This is

$$G(\theta)=\mathbf{e}\exp(-\bar{m})\exp\{\bar{m}H(\theta)\} \qquad \ldots\ldots(2)$$

where $\mathbf{e}$ is the electron charge and $H(\theta)$ is the generating function describing the statistical multiplication process in an electron multiplier containing r uniform dynode stages each of mean gain $\bar{s}$ (Morton and Mitchell 1948, Wright 1954 a).

From this generating function we find that the mean charge arriving at the anode in the interval dt is

$$d\bar{Q}=\left(\frac{\partial G}{\partial\theta}\right)_{\theta=1}=\mathbf{e}\bar{m}\bar{s}^{r}.$$

At a later time t' this charge contributes a charge $d\bar{Q}\exp\{-(t'-t)/RC\}$ to the actual charge on the output capacity at this time. This actual charge is therefore given by

$$\bar{Q}=\int_0^{t'} d\bar{Q}\exp\{-(t'-t)/RC\}=\frac{\bar{Q}_0RC}{\tau-RC}\{\exp(-t'/\tau)-\exp(-t'/RC)\}. \quad ..(3)$$

In this equation $\bar{Q}_0=\mathbf{e}\bar{n}p\bar{s}^{r}$ and is the total charge which flows on to the anode during the whole scintillation.

The variance of the charge which arrives in the interval dt is:

$$\operatorname{var}(dQ)=\left[\frac{\partial^2G}{\partial\theta^2}+\frac{\partial G}{\partial\theta}-\left(\frac{\partial G}{\partial\theta}\right)^2\right]_{\theta=1}=\mathbf{e}^2\bar{s}^{2r}\bar{m}\,\frac{\bar{s}}{\bar{s}-1}. \qquad \ldots.(4)$$

At a later time t' this contributes an amount $\operatorname{var}(dQ)\exp\{-2(t'-t)/RC\}$ to the variance of the actual charge on the capacity at this time. (The factor of 2 in the exponent occurs because the decay constant for the square of the charge is twice that for the charge.)

Thus

$$\left.\begin{aligned}\operatorname{var}(Q)&=\int_0^{t'}\operatorname{var}(dQ)\exp\{-2(t'-t)/RC\}\\ \operatorname{var}(Q)&=\frac{\bar{Q}_0{}^2RC}{2\tau-RC}\,\frac{\bar{s}}{\bar{n}p(\bar{s}-1)}\{\exp(-t'/\tau)-\exp(-2t'/RC)\}.\end{aligned}\right\} \qquad \ldots\ldots(5)$$

Equation (3) describes the mean shape of the scintillation pulse at the multiplier anode and equation (5) describes its irregularity of profile. These equations show that the pulse maximum occurs at a time

$$t_{\mathrm{m}}' = \frac{\tau RC}{\tau - RC} \ln \frac{\tau}{RC}$$

and that the variance is a maximum at a time

$$t_{\mathrm{v}}' = \frac{\tau RC}{2\tau - RC} \ln \frac{2\tau}{RC}.$$

Equations (3) and (5) are illustrated in figure 2 which shows a scintillation pulse for an anthracene crystal. A decay time of 31 mμsec has been used and the mean number of useful photocathode electrons produced during the scintillation has been taken as one hundred. The decay time of the resistance–capacity output circuit has been taken as 6 mμsec; this approximates to the extent of multiplier transit time spread which is of this order in the RCA type 5819 photomultiplier (Hamilton and Wright 1956). Figure 2 (*a*) indicates the irregularity of profile of the typical scintillation pulse and was constructed numerically using a table of random numbers and the expressions for $P(m)$ and $\overline{Q}$. The mean shape of this scintillation pulse is shown by the full curve in figure 2 (*b*); the broken lines in this figure indicate the average fluctuations of the pulse profile from the mean.

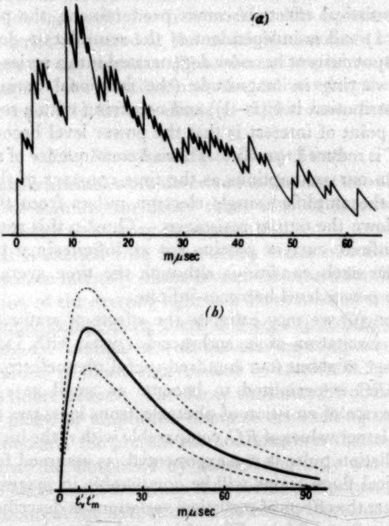


Figure 2. Shape of anthracene scintillation pulse delivered at the multiplier anode. (*a*) schematic illustration of the irregularity of the pulse profile ; (*b*) ———, mean shape of the pulse profile, - - - - - -, average fluctuations of the pulse profile from its mean shape.

Our particular interest is in evaluating the mean power developed by the scintillation pulses. Since the instantaneous output current is proportional to the charge on the output capacity the mean power developed by each pulse is proportional to the integrated mean square charge for each pulse. This may be evaluated immediately from equations (3) and (5). We find:

$$\int_0^\infty \overline{Q^2}\,dt' = \int_0^\infty \{\overline{Q}^2 + \mathrm{var}\,(Q)\}\,dt' = \frac{\overline{Q}_0{}^2R^2C^2}{2}\left[\frac{1}{\tau+RC} + \frac{\bar{s}}{\bar{n}pRC(\bar{s}-1)}\right].$$

The power developed is $W = Z\int\overline{Q^2}\,dt'/RC^2$ where Z is the pulse repetition frequency. Since the mean output current is $i = Z\overline{Q}_0$ we have

$$W = \frac{i^2R}{2Z}\left[\frac{1}{\tau+RC} + \frac{\bar{s}}{\bar{n}pRC(\bar{s}-1)}\right]. \qquad \ldots\ldots(6)$$

Comparing this expression with equation (1) we see that the statistical variations increase the power developed. For small values of RC the additional term describing the statistical effects becomes predominant, the power developed is $i^2R\bar{s}/2Z\bar{n}pRC(\bar{s}-1)$ and is independent of the scintillation decay time τ. This is because the output current has now degenerated into a series of amplified single electron pulses, varying in magnitude (the fractional variance of the single electron pulse distribution is $1/(\bar{s}-1)$) and occurring with a repetition frequency $Z\bar{n}p$. A further point of interest is that the power level becomes infinite as the output capacity C is reduced to zero. This is a consequence of neglecting electron transit times. On our assumptions as the time constant of the anode circuit is reduced to zero the amplified single electron pulses from the multiplier pass instantaneously down the output resistance. Thus in this resistance there is an instantaneously infinite current passing for an infinitesimal time; it is readily verified that under such conditions although the time average of the current remains finite the power level becomes infinite.

Using equation (6) we may estimate the effects of statistical fluctuations in a practical case. Excitation of an anthracene crystal with 5 MeV alpha particles produces an average of about four hundred useful photoelectrons in each scintillation. Even if RC is permitted to become as small as $\tau/10$ the statistical fluctuations in the rate of emission of photoelectrons increases the power level by less than 5%; for larger values of RC, comparable with τ, the increase is negligible. Even if the scintillation pulse is not exponential, as assumed for this discussion, clearly the statistical fluctuations will be comparable in magnitude. In general then we may neglect the effects of statistical variations as described in modification (*a*) and illustrated in figure 2 (*a*). That this may be done is one of the advantages of the present method, for irregularity of the pulse profile has been a major difficulty in previous studies of scintillation decay times. This is particularly so for alpha particles for which it is necessary to study individual scintillations. It is the main reason why no reliable measurements have yet been made of the shapes and durations of alpha particle scintillations.

The effects of variations in pulse size as described in modification (*b*) cannot be neglected, however. In general the scintillation counter produces a broad pulse height distribution. By dividing this distribution into intervals and summing the power developed by pulses in each interval it is evident that the appropriate equation is now

$$W = \left(1 + \frac{\mathrm{var}\,(V)}{\overline{V}^2}\right)\frac{i^2R}{2Z(\tau+RC)} \qquad \ldots\ldots(7)$$

where $\overline{V}$ is the mean of the pulse height distribution and var(V) is its variance. If the pulse height spectrum is very broad (as for gamma ray excitation) it will contain an appreciable number of small pulses for which it may not be justifiable to neglect the statistical variations described in modification (*a*). However, because of their small size such pulses contribute only a very small part towards the total power developed and the error in neglecting these variations and using equation (7) is still very small.

This analysis of stochastic events in the scintillation counter has so far neglected the effects of multiplier dark current, non-uniform dynode sensitivity and multiplier transit time spread. The first two of these are readily included by modifying the expressions for $G(\theta)$ and $H(\theta)$ respectively (Wright 1954 a). They slightly increase the variance of the output charge but the increase in the power output is negligible. It is not possible to make any simple modification of the analysis to include the effects of multiplier transit time spread. However, irrespective of what occurs inside the multiplier it is evident that equation (7) is applicable if the scintillation pulse delivered at the multiplier anode is exponential. It is convenient therefore to use equation (7) to discuss experimental results initially in terms of the shape of the scintillation pulse delivered at the multiplier anode and to correct subsequently for the effects of transit time spread.

2.2. *Photofluorescence Decay Times*

Although the experimental work described in the next two sections is confined to measurements of scintillation decay times it is of interest to discuss how photofluorescence decay times may be measured.

When a crystal is excited with light which varies in intensity in a periodic fashion the fluorescence intensity of the crystal varies in a similar manner. If we express the variation of the exciting light as a Fourier set of components then for each of the harmonic components there is a corresponding component in the fluorescence emission. Compared with the exciting component the fluorescence component is retarded in phase and reduced in amplitude because of the finite decay time of the fluorescence. The decay time may be obtained by measuring the difference in phase or amplitude between the exciting component and the fluorescence component; the present method measures the amplitude difference. The discussion may be simplified by considering only one component of the exciting light. This is justifiable since the various components may be separated with a selective circuit at the output of the detecting instrument, usually a photomultiplier.

The intensity of the exciting light may be expressed as $A+a\sin\omega t$. The intensity of the fluorescence light is then const$\{A+a\sin(\omega t-\phi)/(1+\omega^2t^2)^{1/2}\}$, where τ is the decay time and the constant, less than unity, is introduced because the fluorescence quantum efficiency is less than unity and only a fraction of the fluorescence is collected. The ratio of the fraction modulation of the fluorescence light to the fraction modulation of the exciting light is $D=(1+\omega^2\tau^2)^{-1/2}$. By measuring this ratio the decay time may be obtained. To measure this ratio it is necessary to separate the steady component and the varying component of the multiplier current and measure each separately first for the exciting light and then for the fluorescence light. The former is readily obtained with a suitable meter; the latter can be accurately found by measuring the power developed by it in a resistive load. If the mean current and power are respectively i_E and W_E for the exciting light and are i_F and W_F for the fluorescence light then $D^2 = i_E{}^2 W_F / i_F{}^2 W_E$ and we have $\tau = (1 - D^2)^{1/2} / \omega D$.

The main error in the measurement will be due to multiplier noise arising from statistical fluctuations in the rate of emission of electrons from the photocathode. This may be reduced by using a large photocathode current and a selective circuit with a narrow bandwidth to isolate the varying component.

§ 3. EXPERIMENTAL

The crystals used for these measurements were laminae approximately 1 mm thick freshly cleaved from 1 cm cube crystals obtained for scintillation counting from the Larco Nuclear Instrument Company. The crystal lamina in use was placed on a thin Perspex disc the underside of which was shaped to fit the cathode window of an RCA type 5819 photomultiplier tube mounted vertically. To improve optical contact the lower face of the crystal was polished smooth and a thin film of glycerine was introduced between all surfaces.

For alpha particle measurements a ^{210}Po source was mounted about 1 cm above the crystal. The energy of the particles received by the crystal was 3·6 MeV and these were incident upon a circular area about 2 mm in diameter in the centre of the crystal face; the remainder of the crystal surface was protected by a thin metal foil screen. For electron measurements a ^{137}Cs source was used mounted about 2 cm above the crystal and the whole crystal surface was exposed. Pulse height distribution curves and counting rates were measured using standard scintillation counting techniques. The maximum counting rate used was limited to about 10^4 counts/second so that the correction for counting rate losses should be small.

Approximate calculations showed that under favourable conditions the power developed by the scintillation current would be of the order of a fraction of a microwatt. To measure these very small power levels a thermistor bridge was assembled using Western Electric type D-166382 thermistors. These have an upper frequency limit of 5000 Mc/s which is adequate for scintillation decay time measurements. The experimental arrangement adopted is illustrated in figure 3. Two matched thermistors were placed in series to form one of the arms of a Wheatstone bridge of which all arms had the same resistance 4*R*. To isolate the pulsed scintillation current from the rest of the bridge circuit the thermistors were presented in parallel to the multiplier and formed an effective resistance *R* for the scintillation pulses. The bridge was balanced by adjusting the resistance *r* which varies the bridge current and consequently the resistance of the thermistors.

The thermistors were mounted as near as possible to the multiplier anode pin to reduce stray capacity and inductance; the scintillation pulses were delivered to the thermistors through a non-inductive 500 pF silvered mica capacity. The power developed in the thermistors by the scintillation current changes their temperature and hence their resistance and therefore throws the bridge out of balance. By observing the resulting deflection of the bridge galvanometer a measure of the power developed is obtained. To secure sufficient bridge sensitivity for the measurement of the very small power levels obtainable a lamp and scale galvanometer with a high current sensitivity was used to indicate the bridge balance. Greater sensitivity still was obtained by increasing the scale distance of the galvanometer to four metres. With this arrangement a sensitivity to 10^{-8} watt mm^{-1} was obtained. The large zero drift at this sensitivity was reduced as much as possible by carefully lagging the thermistors with cotton wool. A further reduction in the zero drift was obtained by placing the potentiometer chain of resistances which supplied the multiplier dynode voltages outside the multiplier chassis. This eliminated heating effects from these resistances. With these precautions the zero drift was reduced to about 20 mm/minute.

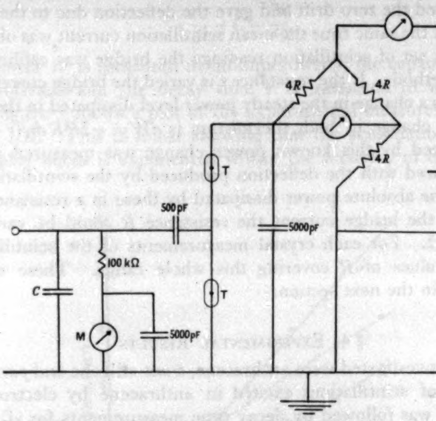


Figure 3. Circuit diagram of the apparatus used for the measurement of scintillation decay times. T, thermistors (Western Electric type D–166382).

To make the power level as large as possible and thus increase the accuracy of measurement in the presence of this zero drift the photomultiplier was operated at the highest gain possible. The maximum usable gain is set by space-charge saturation, which distorts the shape of the pulse, and multiplier stability. The onset of space-charge saturation was investigated for each crystal by measuring the dependence of output power upon output current. At high output currents ($\sim 10\,\mu$A) saturation effects became observable since the power increased less quickly than i^2. The output currents used were therefore kept well below this value. During these measurements it was found that after some little time in operation the multiplier gain became unstable and the dark current increased to several hundred microamperes. Both 5819 tubes available showed this effect even though shielded with an electrostatic screen at cathode potential. The multiplier circuitry was checked for possible feedback effects which might cause this instability (Stump and Talley 1954) but none were found. This gain instability was only eliminated by operating the tube at overall voltages of less than about one thousand volts. At this low voltage dark current and space-charge saturation effects were negligible. Only small scintillation currents could be obtained, however, and the difficulty in measuring the consequent small power levels was the main limitation in the accuracy of these measurements.

With the scintillation pulses delivering power to the thermistors the bridge deflections obtained were about 30 mm for an output current of the order of 1 μA. This deflection was comparable with, but rather larger than, the zero drift which took place during the time required for a measurement. To measure the bridge deflections readings were commenced with the bridge about 100 mm out of balance and drifting towards zero. Readings of the deflection were taken at fifteen second intervals (twice the free period of the galvanometer which was slightly underdamped when in use) with the scintillation current alternately on and off until the bridge was about 100 mm out of balance the other way. Readings were also taken with the zero drift in the opposite direction. Suitably analysed these readings eliminated the zero drift and gave the deflection due to the scintillation power alone. At the same time the mean scintillation current was observed.

Between each set of scintillation readings the bridge was calibrated using a steady current method. If the resistance r is varied the bridge current I is varied and this produces a change in the steady power level dissipated in the thermistor. The total power change in both thermistors is $dW = -2I^2R\,dr/(r+4R)$. The deflection produced by this known power change was measured as described above and compared with the deflection produced by the scintillation pulses in order to obtain the absolute power dissipated by these in a resistance R.

By adjusting the bridge current the resistance R could be varied between 100 Ω and 1000 Ω. For each crystal measurements of the scintillation power were made for values of R covering this whole range. These experimental results are given in the next section.

§ 4. Experimental Results

The crystals investigated were anthracene, *trans*-stilbene and *para*-terphenyl. The decay time of scintillations excited in anthracene by electrons was first measured. This was followed by decay time measurements for all the crystals when excited with alpha particles.

4.1. *Electron Excitation*

Measurements of the power level for scintillations excited in anthracene by electrons are illustrated by curve a of figure 4. This curve is normalized to a mean scintillation current $i = 1\ \mu\mathrm{A}$ and an effective counting rate $Z/(1+\mathrm{var}(V)/\overline{V}^2) = 10^4$ counts/second. The quantity $\mathrm{var}(V)/\overline{V}^2$ was evaluated

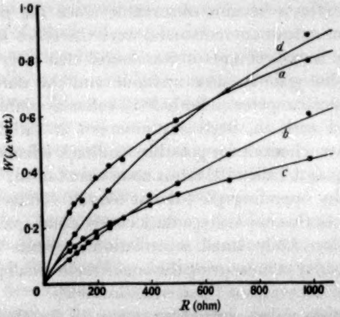


Figure 4. Showing how the power level W, produced by the scintillation current depends upon the resistive load R; (a) electron excitation of anthracene; (b) α-particle excitation of anthracene; (c) α-particle excitation of *trans*-stilbene; (d) α-particle excitation of *para*-terphenyl. Each curve is normalized to a mean scintillation current of 1 μA and an effective counting rate of 10^4 counts/second.

numerically as 0·39 from the pulse height distribution curve measured for the ^{137}Cs gamma rays.

If the scintillation pulses delivered by the multiplier are exponential these measurements may be analysed using equation (7). Inserting the normalized values for the scintillation current and counting rate this equation may be written

$$\frac{\tau}{R}+C=\frac{50}{W}$$

where the power W is measured in millimicrowatts, the capacity C is measured in millimicrofarads and the decay time τ is measured in millimicroseconds. Curve a of figure 5 shows a plot of the experimental measurements giving $1/W$ in terms of $1/R$. This is linear showing that the scintillation pulse delivered *at the multiplier anode* is exponential within the accuracy of the measurements.

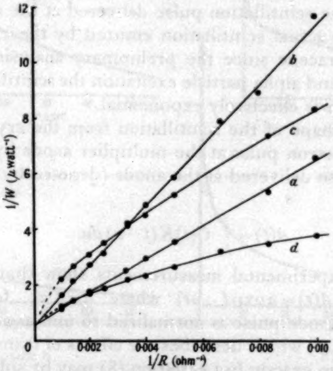


Figure 5. Showing how $1/W$ depends upon $1/R$; (a) electron excitation of anthracene ; (b) α-particle excitation of anthracene ; (c) α-particle excitation of *trans*-stilbene ; (d) α-particle excitation of *para*-terphenyl.

From the slope of this line the decay time was found to be $\tau_e = 31$ mμsec. This agrees with the figure 31 mμsec obtained by Phillips and Swank. From the intercept of this line upon the ordinate the stray capacity of the anode circuit was found to be $C = 29$ pF.

4.2. *Alpha Particle Excitation*

Measurements of the power level for scintillations excited in anthracene, *trans*-stilbene and *para*-terphenyl by alpha particles are illustrated by curves b, c and d respectively of figure 4. These curves are all normalized to a mean scintillation current of 1 μA and an effective counting rate of 10^4 counts/second. The quantity $\mathrm{var}(V)/\bar{V}^2$ was evaluated from the alpha particle pulse height distribution to be less than 0·01 for all three crystals and was consequently neglected.

The scintillations produced in anthracene by alpha particles are found to be exponential *at the multiplier anode*. This is shown by curve b of figure 5 which is linear as for electron excitation. From the slope of this line the decay time was found to be $\tau_a = 53$ mμsec. From its intercept upon the ordinate the stray capacity of the anode circuit was found to be $C = 29$ pF as before.

An analysis of the measurements for *trans*-stilbene and *para*-terphenyl based on equation (7) shows that for each of these crystals the scintillation pulse delivered at the multiplier anode is not exponential. This is shown by curves *c* and *d* of figure 5 both of which are concave downwards instead of being linear as required for an exponential pulse. A more detailed analysis of these measurements cannot be made therefore until more information is available about the shape of the scintillation pulse.

§ 5. DISCUSSION

5.1. *The Crystal Scintillation*

From measurements of the type illustrated in figure 4 it is possible in principle to find the shape of the scintillation pulse delivered at the multiplier anode and then the shape of the actual scintillation emitted by the crystal. This can be done readily for anthracene since the preliminary analysis given above shows that for both electron and alpha particle excitation the scintillation pulse delivered at the multiplier anode is effectively exponential.

If we denote the shape of the scintillation from the crystal by $C(t)$ and the shape of the single electron pulse at the multiplier anode by $K(t)$ then the shape of the scintillation pulse delivered at the anode (denoted by $A(t)$) is given by the integral equation

$$A(t) = \int_0^t C(t) K(t-u)\, du. \qquad \ldots\ldots(8)$$

For anthracene the experimental measurements show that $A(t)$ is exponential and we may write $A(t) = a \exp(-at)$ where $a = 1/\tau$; for convenience this representation of the anode pulse is normalized to unit area. The shape of the single electron pulse $K(t)$ which describes the effects of transit time spread in the multiplier, is not known exactly but equation (8) may be solved for any arbitrary shape of this pulse. In this way it is found that the shape of the scintillation from the crystal is described by the expression:

$$C(t) = \sum_{n=2}^{\infty} \delta^{[n-1]}(t)[a^n c_n - a^{n+1} c_{n+1} + \ldots]/a^{n-1}$$
$$+ \delta(t)[ac_1 - a^2 c_2 + \ldots] + a[1 - ac_1 + a^2 c_2 - \ldots]\mathrm{e}^{-at}. \qquad \ldots\ldots(9)$$

The coefficients $c_1, c_2, \ldots$ are defined by the equation

$$p/k(p) = 1 + c_1 p + c_2 p^2 + c_3 p^3 + \ldots$$

where
$$k(p) = p \int_0^{\infty} \mathrm{e}^{-pt} K(t)\, dt$$

and $K(t)$ has been normalized to unit area.

The first term in this expression contains impulse functions $\delta^{[n]}(t)$ of order higher than the first and it does not contribute to the scintillation since its integral is zero. Its presence is required to produce a mathematically exact solution to equation (8). The second term contains the impulse function of the first order and describes an initial spike of fluorescence. The third term describes an exponential fluorescence component with the same decay time as that observed at the multiplier anode.

For both electron and alpha particle excitation of anthracene then the scintillation from the crystal consists of an initial surge of fluorescence followed by an exponential tail. We may estimate the proportion of fluorescence contained in the initial surge by using the expression $K(t)=b^2t\exp(-bt)$ to describe the shape of the single electron pulse. This expression has been used previously and gives a good representation of the single electron pulse (Hamilton and Wright 1956); its shape is shown in figure 6 and may be compared with the shape shown in figure 7 which was measured for the C–7165A multiplier (Swank 1954).

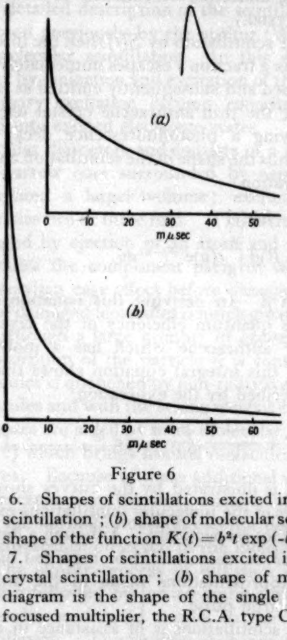


Figure 6

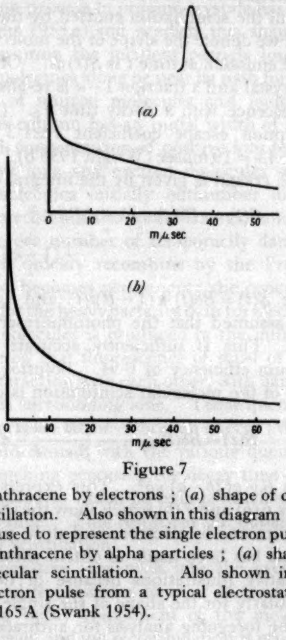


Figure 7

Figure 6. Shapes of scintillations excited in anthracene by electrons ; (*a*) shape of crystal scintillation ; (*b*) shape of molecular scintillation. Also shown in this diagram is the shape of the function $K(t)=b^2t\exp(-bt)$ used to represent the single electron pulse.

Figure 7. Shapes of scintillations excited in anthracene by alpha particles ; (*a*) shape of crystal scintillation ; (*b*) shape of molecular scintillation. Also shown in this diagram is the shape of the single electron pulse from a typical electrostatically focused multiplier, the R.C.A. type C–7165A (Swank 1954).

Retaining only the physically real parts of the expression for $C(t)$ we find:

$$C(t)=\frac{a}{b^2}(2b-a)\delta(t)+\frac{(b-a)^2}{b^2}\,ae^{-at}. \qquad \ldots\ldots(10)$$

It is readily verified that $C(t)$ has unit area as expected. For the RCA type 5819 photomultiplier the width at half maximum of the single electron pulse is about 6 mμsec giving $b=4\times10^8\ \text{sec}^{-1}$. Using the appropriate measured values for the scintillation decay times we find that for electron excitation a fraction of about 0·15 of the total scintillation is contained in the initial spike and for alpha particle excitation a fraction of about 0·09 is contained in this spike. (These figures are only approximate because the shape and duration of the single electron pulse are known only approximately.) The shape of the scintillation emitted by the crystal is illustrated in figure 6 (*a*) for electron excitation and in figure 7 (*a*) for alpha particle excitation. In these diagrams the initial spike of fluorescence has been represented schematically.

5.2. *The Molecular Scintillation*

Many organic crystals are opaque to part of their own fluorescence and self-absorption of fluorescence in these crystals modifies the shape of the scintillation emitted. Self-absorption is not a fundamental part of the scintillation process and it is necessary to estimate its effects if the true shape of the scintillation is to be determined. (This may be termed the molecular scintillation to distinguish it from the scintillation emitted by the crystal.)

If we denote the shape of the molecular scintillation by $S(t)$ then the intensity of the emission at time t is $S(t)\,dt$. Of this a fraction κ escapes immediately from the crystal and a fraction $1-\kappa$ is re-absorbed and subsequently emitted as photofluorescence with a decay time τ_p. (For the thin anthracene crystal used the absorption escape coefficient $\kappa \simeq 1/3$ giving a photofluorescence decay time $\tau_p = 6{\cdot}4/\kappa = 19$ mμsec (Wright 1955 b).) Thus the shape of the scintillation emitted by the crystal is given by the integral equation:

$$C(t)=\kappa B\delta(t)+(1-\kappa)B\beta e^{-\beta t}+\kappa(1-B)s(t) + (1-\kappa)(1-B)\beta\int_0^t s(u)e^{-\beta(t-u)}\,du \qquad \ldots\ldots(11)$$

where $S(t)=B\delta(t)+(1-B)s(t)$ and $\tau_p=1/\beta$. In deriving this equation it has been assumed that the photofluorescence quantum efficiency of the crystal is unity. This is sufficiently accurate for anthracene which has a molecular quantum efficiency of 0·94. Solution of this integral equation shows that the shape of the molecular scintillation is described by the expression

$$S(t)=B\delta(t)-\frac{1-\kappa}{\kappa}\,\frac{\beta B-a}{\beta/\kappa-a}\,\frac{\beta}{\kappa}\,e^{-\beta t/\kappa}+\frac{1-\kappa B}{\kappa}\,\frac{\beta-a}{\beta/\kappa-a}\,a\,e^{-at} \qquad \ldots\ldots(12)$$

where $B=a(2b-a)/\kappa b^2$. This expression is illustrated by the curves shown in figures 6 (*b*) and 7 (*b*) which show the shapes of the molecular scintillations excited in anthracene by electrons and alpha particles respectively. Although a detailed analysis has been made to obtain these curves, they should be regarded only as qualitative illustrations because of the approximations made in the analysis particularly for the shape of the single electron pulse.

The foregoing analysis for anthracene scintillations is of assistance in interpreting the measurements for *trans*-stilbene and *para*-terphenyl for we may expect all the organic crystal phosphors to have similar characteristics. Numerical calculations show that if the scintillation pulse from the multiplier has a leading edge which is appreciably peaked then a plot of the type shown in figure 5 is concave downwards; if the leading edge of the pulse is appreciably depressed then this plot is concave upwards. The experimental curves plotted in figure 5 for these two crystals are decidedly concave downwards showing that the scintillation pulses delivered *at the multiplier anode* have leading edges which are appreciably peaked. In view of the above analysis for anthracene these results are interpreted as showing that the scintillations emitted by these crystals commence with an initial surge of fluorescence which contains an appreciable proportion of the whole scintillation emission. For both these crystals the initial spike is larger than for anthracene since it is apparent even at the multiplier anode. This is because self-absorption in *trans*-stilbene is small and in *para*-terphenyl is negligible (Wright 1954 d, Birks and Wright 1954). The shape of the molecular scintillation is therefore little affected by self-absorption and the initial spike of fluorescence, which in anthracene is much reduced by this effect, is only slightly diminished in the scintillations from these crystals.

The foregoing analysis of the experimental measurements shows that for all three crystals investigated the molecular scintillation consists of an initial surge or spike of fluorescence followed by a component of much longer duration and non-exponential shape; most of the emission is contained in the long component of the scintillation.

A detailed description of the scintillation process in organic crystals has been proposed previously by the author (Wright 1955 a) and predicts this shape for the molecular scintillation. In this description the incident particle loses its energy by ionization and excitation of the molecules along or near its path forming a primary excitation column consisting of ionized molecules and molecules raised into excited electronic states. This column has an initial width of a few molecular diameters and consists of a high concentration of positive ions located in a narrow core surrounded by negative ions and excited molecules spread throughout a larger volume; excited molecules initially outnumber ionized molecules two or three fold. The excitation column also contains some molecules damaged by ejection of an atom and a larger number of temporarily damaged molecules the component parts of which quickly recombine by the Franck–Rabinowitch cage effect before dissociation becomes permanent; the concentration of damaged molecules is much greater for the heavy particles than for electrons.

Although a large number of excited molecules is produced in the initial act of absorption of the particle it is considered that fluorescence of most of these molecules is quenched by non-radiative interaction with each other, with damaged molecules and with the strong electric fields surrounding ions. These quenching processes are aided by intra-molecular migration of the excitation energy (Wright 1955 c) which brings excited molecules into contact with the various quenching centres. Because of these additional quenching processes the decay time of the fluorescence from the small number of molecules which do emit is much shorter than the normal photofluorescence decay time. This results in the initial spike of fluorescence observed in the molecular scintillation.

The slower process of diffusion of negative ions back into the excitation column followed by ion recombination and fluorescence of most of the excited molecules so formed results in the long decay component of the molecular scintillation. This process of charge diffusion and ion recombination is produced by the steep potential gradients existing in the region of the excitation column and becomes slower as these potential gradients are reduced. It is possible that weak emission during the final stages of this diffusion process is partly responsible for the very long decay components (of the order of a few microseconds) observed in scintillations from organic crystals (Harrison 1954).

For scintillations produced by alpha particle excitation of anthracene it is observed that the duration of the scintillation is longer than for electron excitation. This is a consequence of the greater excitation densities produced by the heavy particles. Because of these high excitation densities most of the excited molecules formed in the initial stages of ion diffusion and recombination are quenched by the high concentration of ions still remaining. This reduces fluorescence emission during the early stages of the scintillation and results in an effective increase in the observed duration of the scintillation.

ACKNOWLEDGMENTS

The author wishes to thank the South African Council for Scientific and Industrial Research for a research grant to purchase equipment and the Telecommunications Research Laboratory, Johannesburg, for the loan of the thermistors used.

REFERENCES

BAILEY, E. A., and ROLLEFSON, G. K., 1953, *J. Chem. Phys.*, **21,** 1315.
BIRKS, J. B., 1953, *Scintillation Counters* (London: Pergamon Press; New York; McGraw-Hill).
BIRKS, J. B., and LITTLE, W. A., 1953, *Proc. Phys. Soc.* A, **66,** 921.
BIRKS, J. B., and WRIGHT, G. T., 1954, *Proc. Phys. Soc.* B, **67,** 657.
CURRAN, S. C., 1953, *Luminescence and the Scintillation Counter* (London: Butterworths Scientific Publications).
GARLICK, G. F. J., and WRIGHT, G. T., 1952, *Proc. Phys. Soc.* B, **65,** 415.
HAMILTON, T. D. S., and WRIGHT, G. T., 1956, *J. Sci. Instrum.*, **33,** 36.
HARRISON, F. B., 1954, *Nucleonics*, **12,** 24.
LIEBSON, S. H., 1952, *Nucleonics*, **7,** 41.
MORTON, G. A., and MITCHELL, J. A., 1948, *R.C.A. Review*, **9,** 632.
PHILLIPS, R. B., and SWANK, R. K., 1953, *Rev. Sci. Instrum.*, **24,** 611.
REYNOLDS, G. T., 1952, *Nucleonics*, **7,** 46.
STUMP, R., and TALLEY, H. E., 1954, *Rev. Sci. Instrum.*, **25,** 1132.
SWANK, R. K., 1954, *Nucleonics*, **12,** 14.
WRIGHT, G. T., 1953, *Phys. Rev.*, **91,** 1282; 1954 a, *J. Sci. Instrum.*, **31,** 377; 1954 b, *Ibid.*, **31,** 462; 1954 c, *Phys. Rev.*, **96,** 569; 1954 d, S. A. Congress of Physics, Pretoria, July 1954; 1955 a, *Proc. Phys. Soc.* B, **68,** 929; 1955 b, *Ibid.*, **68,** 241; 1955 c, *Ibid.*, **68,** 701.
WRIGHT, G. T., and GARLICK, G. F. J., 1954, *Brit. J. Appl. Phys.*, **5,** 13.

Two-dimensional Strain Measurement by Moiré

BY R. H. BROMLEY
Printing, Packaging and Allied Trades Research Association, Leatherhead, Surrey

MS. received 22nd April 1955 *and in amended form 22nd September* 1955

Abstract. Applications of moiré to surface strain measurement are described. A theory of moiré is developed, based on the concept of the covariant vector, which leads to a simple description of infinitesimal strain fields. Extension to the case of finite strain is considered, and parallels with crystallography are drawn.

§ 1. INTRODUCTION

IN the investigation of the rheological properties of paper, fabrics and other materials in sheet form it is necessary to determine their dimensional changes whilst being subjected to stress or moisture absorption for example. Precise determination of these changes is not easy and the methods at present in use are generally confined to an overall measurement in one direction (e.g. the principal direction of stress in a tensile test). The technique outlined below yields accurate information on the complete strain tensor of two-dimensional homogeneous deformation and provides sufficient data for the calculation of the whole strain field in the heterogeneous case.†

§ 2. PATTERNS AND MOIRÉ

Patterns are a feature of many sheet materials as, for example, the weave of fabrics or the follicles of leather, and a plain surface such as that of paper can be provided with a simple pattern by printing a uniform half-tone upon it. The method of strain measurement to be described requires the presence of such a pattern which, in so far as it is uniform, consists of a basic unit (the 'motif') repeated at regular intervals in all directions. The actual form of the motif is relatively unimportant and it is convenient to abstract from the pattern a regular array of points, which we call its *lattice*, such that the points may all be made to correspond with a given element of the motif (Weyl 1952).

An interesting phenomenon occurs when two or more patterns are superimposed in any way. The mutual 'interference' results in the emergence of a new pattern which is known as a *moiré*. This phenomenon arises in the printing industry in connection with multi-colour printing and the reproduction of half-tones by re-screening (Tollenaar 1945) and the examples of moiré given in figure 1 (Plate) have been obtained with printing screens. These examples illustrate the characteristic appearance of moiré, that is, one or more sets of 'interference' fringes.

---

† Earlier work on rheological applications of moiré has recently come to the notice of the author. This is described by Kaczér and Kroupa (1952) and Ligtenberg (1952).

§ 3. THEORY OF MOIRÉ

Consider first the simple case of two line screens approximately equal in size superimposed at a small angle (figure 1(*a*)). The resultant density of the moiré is least at the points where the two sets of lines intersect. Accordingly the fringes, which are regions of maximum density, run between the rows of intersections as shown in figure 2 and the fringe separation can be calculated by simple trigonometry from the two screen sizes and the angle of superposition.

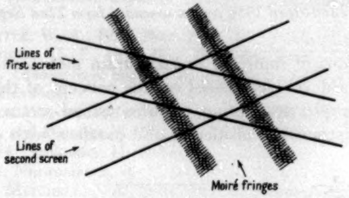


Figure 2. Construction of moiré of two line screens.

This diagram immediately suggests an analytical formulation in terms of covariant vectors, which greatly facilitates the interpretation of more complex moiré patterns. A two-dimensional covariant vector, w_λ say, is a tensor quantity which can be represented geometrically by an ordered pair of parallel lines, the components w_1 and w_2 being the reciprocals of the intercepts on the coordinate axes. (In three dimensions a covariant vector would be represented by a pair of parallel planes, and so on.) This is analogous to the representation of an ordinary (contravariant) vector by an ordered pair of points† and in contradistinction to the 'parallelogram law' of addition, covariant vectors add according to the scheme shown in figure 3 (Schouten 1951). The sense of the vectors has been indicated by an arrow (not necessarily straight) joining each pair of lines. The similarity between figures 2 and 3 is evident, and enables us to say that if the two original line screens represent (approximately equal) covariant vectors the moiré fringes represent their covariant vector difference.

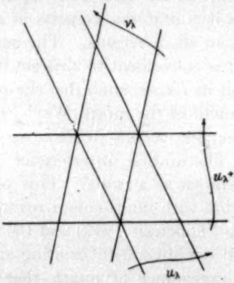


Figure 3. Addition of two-dimensional covariant vectors.

---

† It is true that a contravariant vector is commonly represented as an arrow. However, the barb is simply an ordering device and the line need not be straight, only the end-points being fixed relative to one another.

To generalize this result we must return to the array of points which constitute the lattice of a pattern. This is the two-dimensional counterpart of a crystal lattice and corresponding to the crystallographic 'planes' we have sets of parallel lines passing through the points in various directions as illustrated in figure 4. For want of a better term we shall call any one set of lines a lattice ***stratum*** and, clearly, with each stratum a covariant vector can be associated. A coordinate system of reference which has the two greatest[†] lattice strata for its covariant base vectors we shall call a ***cardinal*** coordinate system for the lattice. Relative to such a system of reference the components of a stratum are always a pair of relatively prime integers, corresponding to the Miller indices of a crystallographic 'plane'. Any pair of lattice strata with the same unit area as the base vectors of the cardinal coordinate system may be said to generate the given lattice.

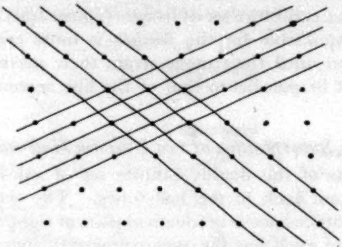


Figure 4. Examples of lattice strata.

Each stratum within a pattern has its characteristic size and contrast, the former depending solely on the lattice structure[†]. The contrast of a stratum w_λ may be defined as the range of mean density observed through a thin slit which is translated over w_λ and this obviously depends on the motif. As the largest lattice stratum in a printed half-tone also has the highest contrast there can be little doubt of its pre-eminence in the structure of this pattern: the situation is not so simple when two or more half-tones are superimposed because the strata of the separate patterns are supplemented by others arising through the 'interference'. In fact, every pair of independent strata must be imagined to generate a 'sub-lattice', the 'sub-strata' of which all contribute to the composite pattern. The proportionality between size and contrast does not necessarily apply to these sub-strata: in general they are of low contrast, but in certain cases and for particular pairs of strata they may be very large, so that they become the most distinct elements of the pattern. Evaluation of the appearance of a moiré, depending as it does on the relative size and contrast of its various strata, can be extremely complex. We shall discuss by way of illustration the examples of moiré already given.

3.1. *Superposition of two Line Screens*

A line screen is really a one-dimensional pattern, its lattice having only one stratum of non-zero magnitude, and accordingly the moiré produced by the interaction of two line screens contains only one sub-lattice. In favourable

---

[†] The magnitude $|w|$ of a covariant vector w_λ is an invariant defined by

$$\frac{1}{|w|^2} \overset{\text{def}}{=} g^{\kappa\lambda} w_\kappa w_\lambda .$$

cases the largest stratum of this sub-lattice appears as a uniform system of fringes, *a fortiori* because it also has the highest contrast in this simple type of moiré.

An important special case is that in which the finite strata of the two screens are roughly equal (in magnitude and direction). The sub-stratum given by their covariant vector difference is then much larger than any other, and produces a set of fringes (figure 1(*a*)). If the relative orientation of the two screens is varied, their strata are most nearly equal when they are parallel, and at this point the fringes are of maximum size.

3.2. *Superposition of a Line Screen and a Half-tone*

Each sub-lattice in the moiré resulting from this superposition arises from the interaction of the finite stratum of the line screen and a different stratum of the half-tone. If one sub-lattice contains a stratum which is far more prominent than any other the moiré exhibits a set of fringes (figure 1(*b*)). The stratum of the half-tone which is responsible for the fringes is most easily distinguished by rotating the line screen until the fringes attain their maximum size, when the required stratum must be parallel to that of the line screen, and this suffices to determine it uniquely.

3.3. *Superposition of two Uniform Half-tones*

This moiré consists of the doubly infinite set of sub-lattices generated by pairs of strata, one from each of the half-tones. The symmetry of the latter demands that the sub-lattices occur in identical pairs at right angles to one another, and therefore the moiré itself has the same symmetry (figure 1(*c*)). In general it contains two prominent sets of fringes of equal size in perpendicular directions, and the strata whose interference produces these fringes may be found by trial and error. In the particular case of two half-tones of equal size, however, there are just two possibilities. When the two patterns are not far out of coincidence (i.e. an angular deviation of less than 30°) the interfering strata belong to the form $\{1, 0\}$,[†] producing what Tollenaar calls the primary moiré. Between 30° and 45° the so-called secondary moiré appears, in which the fringes arise from the interaction of strata belonging to the form $\{2, 1\}$.

§ 4. MEASUREMENT OF INFINITESIMAL STRAIN IN TWO DIMENSIONS[‡]

If there is not already a pattern on the surface a dark half-tone is printed upon it preferably when the material is in its standard (strain-free!) condition. The orientation of the print with respect to the surface is not important, although a special choice may simplify the subsequent calculation. To fix ideas we shall assume the pattern to be such a half-tone: the modifications necessary to utilize an existing pattern should be quite evident.

When the material is ready for testing, the pattern is photographed, and after the deformation has occurred it is photographed a second time on the same plate. Now consider the stratum w_λ at the point ξ^κ of the half-tone. On the photographic plate the strain-displaced vector from the second exposure is superposed on a parallel-displaced vector from the first (figure 5). Clearly the major contribution to the resulting moiré is their covariant vector difference. It can be shown (see Appendix) that this difference is given by

---

[†] The term ' form ' is used here in its crystallographic sense.

[‡] We shall make use of the ' kernel-index ' tensor notation described by Schouten, in which each quantity is denoted by a fixed kernel letter, transformation to another coordinate system being indicated by a change in the type of index (from unprimed to primed indices, for example). This accords more nearly with physical ideas than does the usual notation.

$$\delta w_\lambda = -w_\sigma u^\sigma{}_{,\lambda} \qquad \ldots\ldots(1)$$

where u^κ is the (small) displacement of the material at the point ξ^κ and the comma denotes covariant differentiation. The moiré contains two fringe systems, which are obtained from (1) by taking for w_λ the covariant base vectors of the cardinal coordinate system, say (κ). Denoting the base vectors by $\overset{\kappa}{e}_\lambda$ these moiré strata become $\delta\overset{\kappa}{e}_\lambda$, and we have from equation (1)

$$\begin{aligned} \delta\overset{\kappa}{e}_\lambda &= -\overset{\kappa}{e}_\sigma u^\sigma_\lambda \\ &\overset{*}{=} -u^\kappa{}_{,\lambda} \\ &\overset{*}{=} -u_{\kappa,\lambda} \text{ since } (\kappa) \text{ is here Cartesian.}\dagger \end{aligned}$$

Hence with the usual definition of infinitesimal strain,

$$\begin{aligned} \epsilon_{\kappa\lambda} &\overset{\text{def}}{=} \tfrac{1}{2}(u_{\kappa,\lambda} + u_{\lambda,\kappa}) \\ &\overset{*}{=} -\tfrac{1}{2}(\delta\overset{\kappa}{e}_\lambda + \delta\overset{\lambda}{e}_\kappa). \qquad \ldots\ldots(2) \end{aligned}$$

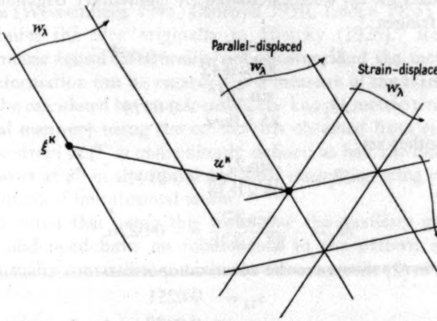


Figure 5. Construction of moiré of infinitesimal strain.

In practice it is not necessary to use photography. The original pattern can be preserved by printing the half-tone on to a thin sheet of glass or some other transparent material. This is then superposed on the strained material with the two patterns as nearly as possible in register to obtain the moiré. The effect of any small disorientation is eliminated during the calculation of the strain tensor by virtue of the mixing process in its definition.

The moiré obtained in this way can be improved in contrast by using a light tone for the transparent print, roughly complementary to the one on the surface. For maximum contrast the latter should be of uniform 54% printing area and the transparency made from it in the form of an equal-size photographic negative.

§ 5. EXAMPLE OF HOMOGENEOUS STRAIN

The method outlined above was used to measure the swelling of flong (a hygroscopic papier-maché material used in stereotyping). A uniform 60% printing area half-tone with 85 dots to the inch was printed on an air-dry sample

---

† The sign $\overset{*}{=}$ means that the equation is not in invariant form and is not necessarily valid for coordinate systems other than those specified.

of the flong containing 5% moisture. The transparent grid was made by preparing a glass negative of the same size and marking on it a pair of cardinal axes in indian ink.

The flong was then damped down to a moisture content of 53% and the transparent grid placed on top with the two printed surfaces in contact and approximately in register. Before proceeding to measure the moiré fringes it was necessary to label them correctly. The vectors $\delta\overset{1}{e}_{\lambda}$ and $\delta\overset{2}{e}_{\lambda}$ were readily distinguished since $\delta\overset{\kappa}{e}_{\lambda}$ was either parallel to e^{λ}_{κ} or could be made so by a small rotation of the transparent grid. The sense of $\delta\overset{\kappa}{e}_{\lambda}$ was then the same as that of $\overset{\kappa}{e}_{\lambda}$ if it rotated through this position in the opposite direction to that of the transparent grid and vice versa.

Having identified the $\delta\overset{\kappa}{e}_{\lambda}$ with the moiré fringes the transparent grid was carefully adjusted to make them as large as possible. They were then measured by means of a Douglas protractor which enabled the magnitude and direction of a stratum to be measured simultaneously. The moiré and its fringe measurements are shown in figure 6 (Plate).

From these data the $\delta\overset{\kappa}{e}_{\lambda}$ were calculated by elementary trigonometry. Thus, from one set of fringes:

$$\delta\overset{1}{e}_{1}=-\frac{\cos 26\tfrac{1}{4}^{\circ}}{85\times 0{\cdot}42}=-0{\cdot}0251$$

and

$$\delta\overset{1}{e}_{2}=\frac{\sin 26\tfrac{1}{4}^{\circ}}{85\times 0{\cdot}42}=0{\cdot}0124$$

whilst from the other set:

$$\delta\overset{2}{e}_{1}=\frac{\sin 27^{\circ}}{85\times 0{\cdot}38}=0{\cdot}0141$$

and

$$\delta\overset{2}{e}_{2}=-\frac{\cos 27^{\circ}}{85\times 0{\cdot}38}=-0{\cdot}0276.$$

Substitution in (2) then gave the strain components

$$\begin{aligned}\epsilon_{11}&=0{\cdot}0251\\ \epsilon_{12}=\epsilon_{21}&=-0{\cdot}0132\\ \epsilon_{22}&=0{\cdot}0276.\end{aligned}$$

This tensor has principal components 0·013 and 0·040, found by rotating the (κ)-axes through $42\tfrac{1}{2}^{\circ}$, giving the strain in the machine and cross directions respectively.

§ 6. MEASUREMENT OF HETEROGENEOUS STRAIN

The method of strain measurement described above may not be suitable in the case of small heterogeneous strains because the moiré fringe pattern is a large one, making it impossible to study small areas separately. The difficulty can be overcome by a suitable rotation of the auxiliary grid relative to the surface, making the moiré fringes as small as desired, but this renders the foregoing analysis no longer applicable since it can only allow for infinitesimal rotations. The grid and surface patterns must then be kept distinct, each having its own cardinal coordinate system, which we denote by (κ) and (κ') respectively. Provided the strain is not too large (but possibly finite, nevertheless) the moiré fringes measure the differences between the covariant base vectors of (κ) and (κ') at each point, e.g. using the same notation as far as possible,

$$\delta^{\kappa}_{\cdot\lambda}=\overset{\kappa'}{e}_{\lambda}-\overset{\kappa}{e}_{\lambda} \qquad \ldots\ldots(3)$$

whence the base vectors of (κ') can be determined relative to (κ). An estimate of the angle by which the patterns have been rotated out of register can then be found: it is, in fact, approximately $\tan^{-1}\{(\overset{2'}{e}_1 - \overset{1'}{e}_2)/(\overset{1'}{e}_1 + \overset{2'}{e}_2)\}$ at each point. Let (κ'') be the coordinate system obtained by rotating (κ) through such an angle. The base vectors of (κ') can be transformed to this system of reference and then substituted back into equation (3), re-written with respect to (κ''). This coordinate system is sufficiently coincident with that on the material for the components $\delta\overset{\kappa''}{e}_{\lambda''}$ obtained to be substituted into equation (2) to give the strain tensor relative to (κ').

Alternatively, if the strain must be considered finite we may proceed as follows. From the moiré fringes the $\overset{\kappa'}{e}_\lambda$ are calculated as before. The transformation $\xi^\kappa \rightarrow \xi^{\kappa'}$ is then known, since (Schouten 1951)

$$A^{\kappa'}_{\lambda} \overset{\text{def}}{=} \frac{\partial \xi^{\kappa'}}{\partial \xi^{\lambda}} \overset{*}{=} \overset{\kappa'}{e}_\lambda . \qquad \text{......(4)}$$

The coordinate system (κ') deforms with the surface, since it is embedded in it, and the use of such a 'convected' system of reference has been advocated by several authors (Weissenberg 1948, Oldroyd 1950, Lodge 1953) amongst whom Oldroyd attributes the idea originally to Hencky (1925). Referred to such a system the metric tensor is naturally not constant and the metric field at any stage of the deformation can be regarded as a measure of the form of the surface. This field can be calculated by transforming the known metric tensor of Cartesian (κ) in the usual manner† using the coefficients obtained from equation (4).

Finally, the strain at $\xi^{\kappa'}$ is conveniently defined as half the difference between the metric tensors at $\xi^{\kappa'}$ in the initial and final stages, agreeing in the limit with the usual definition of infinitesimal strain.

It is to be noted that using this technique the auxiliary grid is purely an intermediary, and need have no relationship to the pattern on the material, although it is usually convenient to make the two congruent for some condition of the material.

§ 7. EXAMPLE OF FINITE HETEROGENEOUS STRAIN

A specimen of flong 2 inches wide, carrying the same 85-screen half-tone, was stretched at the rate of 1 in. per minute on a Hounsfield tensometer with the clamps initially 6 in. apart. Shortly before the breaking strain was reached the machine was stopped and the specimen removed. The non-uniform permanent strain remaining in it gave rise to the rather irregular moiré shown in figure 7 (Plate) when the transparent grid was placed over it. Three points, marked A, B and C, were chosen to illustrate the finite theory propounded above. The procedure for measuring the moiré fringes remained unchanged except that care was necessary to make the measurements exactly at the points in question.

At point A the moiré components were found to be‡

$$\delta\overset{\kappa'}{e}_\lambda = \begin{pmatrix} -0{\cdot}0356 & -0{\cdot}0444 \\ 0{\cdot}1063 & -0{\cdot}0207 \end{pmatrix}$$

and so, from (3),

$$\overset{\kappa'}{e}_\lambda = \begin{pmatrix} 0{\cdot}9644 & -0{\cdot}0444 \\ 0{\cdot}1063 & 0{\cdot}9793 \end{pmatrix}.$$

---

† We have $g_{\kappa'\lambda'} = A^{\kappa}_{\kappa'} A^{\lambda}_{\lambda'} g_{\kappa\lambda}$ where $A^{\kappa}_{\kappa'} A^{\kappa'}_{\lambda} \overset{*}{=} \delta^{\kappa}_{\lambda}$ (the Kronecker delta).

‡ In dealing with expressions with two indices and their corresponding matrices we agree that the first (or upper) index numbers the columns and the second (or lower) the rows.

Following the procedure outlined above, the metric tensor of deformed (κ') was found to be

$$g_{\kappa'\lambda'}=\begin{pmatrix}1\cdot0666 & -0\cdot0680\\ -0\cdot0680 & 1\cdot0448\end{pmatrix},$$

and the strain from the initial state

$$\epsilon_{\kappa'\lambda'}=\begin{pmatrix}0\cdot0333 & -0\cdot0340\\ -0\cdot0340 & 0\cdot0224\end{pmatrix}.$$

Similarly the moiré components at B and C were

$$\begin{pmatrix}-0\cdot0309 & -0\cdot0520\\ 0\cdot1077 & -0\cdot0169\end{pmatrix} \quad\text{and}\quad \begin{pmatrix}-0\cdot0450 & -0\cdot0193\\ 0\cdot1087 & -0\cdot0341\end{pmatrix};$$

which gave for the strain at these points, relative to (κ'),

$$\begin{pmatrix}0\cdot0277 & -0\cdot0302\\ -0\cdot0302 & 0\cdot0177\end{pmatrix} \quad\text{and}\quad \begin{pmatrix}0\cdot0484 & -0\cdot0508\\ -0\cdot0508 & 0\cdot0428\end{pmatrix} \text{ respectively.}$$

The orientation of the printed half-tone was such that the axes of (κ') were initially inclined to the edges of the specimen at 45°. The strain components relative to axes parallel to these edges were therefore

$$\begin{pmatrix}0\cdot062 & -0\cdot005\\ -0\cdot005 & -0\cdot006\end{pmatrix} \begin{pmatrix}0\cdot053 & -0\cdot005\\ -0\cdot005 & -0\cdot008\end{pmatrix} \begin{pmatrix}0\cdot096 & -0\cdot003\\ -0\cdot003 & -0\cdot005\end{pmatrix}$$

at the points A, B and C respectively.

§ 8. COMPARISON WITH X-RAY CRYSTALLOGRAPHY

The correspondence between pattern and crystal extends to the experimental techniques used in studying them. Thus, the investigation of crystals by means of plane x-radiation is analogous to obtaining a moiré between a pattern and a line screen in the manner of figure 1(*b*). The typical moiré fringes cannot be observed in the former case, however, because their wavelength is so large as to give rise to complete absorption. The Bragg equation expresses a sufficient condition that a stratum exists in the three-dimensional moiré with the same wavelength as the incident x-rays. It is therefore not absorbed, but appears as 'reflected' radiation. This would seem to be a primitive application of the moiré technique which is dictated by the physical conditions of the experiment.

§ 9. CONCLUSIONS

Methods have been described which are capable of giving full information about two-dimensional strain fields with a minimum of apparatus. Where the strain is supposed to be homogeneous a direct check is available in the uniformity of the moiré pattern and the technique should have wide applications in the study of heterogeneous strain fields.

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. W H. Banks for suggesting the work and to acknowledge the helpful criticism and advice given by Dr. K. Weissenberg, Mr. G. A. Gordon, and by the Physical Society's referee concerning the presentation. Thanks are also due o Mr. L. E. Boxall and Mr. G. A. Mitchener for technical assistance and to the Council and Director of the Printing, Packaging and Allied Trades Research Association for permission to publish this paper.

REFERENCES

HENCKY, H., 1925, *Z. angew. Math. Mech.*, **5**, 144.

KACZÉR, J., and KROUPA, F., 1952, *Czech. J. Phys.*, **1**, 80.

LIGTENBERG, F. K., 1952, *De Ingenieur, s'Grav*, **64** (No. 9), O 42.

LODGE, A. S., 1953, *2nd Int. Congr. Rheol.*

OLDROYD, J. G., 1950, *Proc. Roy. Soc.*, **200**, 523.

SCHOUTEN, J. A., 1951, *Tensor Analysis for Physicists* (Oxford : University Press).

TOLLENAAR, D., 1945, *Moiré, Interferentieverschijnselen bij rasterdruck* (Amsterdam: Instituut voor Grafische Techniek).

WEISSENBERG, K., 1948, *1st. Int. Congr. Rheol.*

WEYL, H., 1952, *Symmetry* (Princeton : University Press).

APPENDIX

DERIVATION OF EQUATION (1)

Following Schouten, let the particles of a piece of matter which occupies a region of ordinary space undergo a small displacement u^κ, and let w_λ be an infinitesimal vector at the point ξ^κ which is fixed in the matter. Then if (κ') is the coordinate system arising from (κ) by dragging along over the infinitesimal vector u^κ we know that (Schouten 1951)

$$A^{\kappa'}_{\lambda} \overset{*}{=} \delta^{\kappa'}_{\kappa}\left(A^{\kappa}_{\lambda} - \frac{\partial u^\kappa}{\partial \xi^\lambda}\right),$$

and consequently the displaced vector w_λ in $\xi^\kappa + u^\kappa$ has the components

$$w_\lambda - w_\kappa \frac{\partial u^\kappa}{\partial \xi^\lambda} \text{ with respect to } (\kappa).$$

If the vector w_λ is parallel-displaced to $\xi^\kappa + d\xi^\kappa$ the components become

$$w_\lambda + \Gamma^{\kappa}_{\sigma\lambda} w_\kappa u^\sigma.$$

Hence the variation of the vector w_λ seen from the point of view of an observer whose system of reference undergoes parallel displacement is

$$\begin{aligned} \delta w_\lambda &= -w_\kappa \frac{\partial u^\kappa}{\partial \xi^\lambda} - \Gamma^{\kappa}_{\sigma\lambda} w_\kappa u^\sigma \\ &= -w_\sigma u^\sigma{}_{,\lambda}. \end{aligned}$$

This is the quantity which, in the particular case of two-dimensional deformation, can be obtained from the moiré pattern.

The Propagation and Reflection of Sound Pulses of Finite Amplitude

By T. F. W. EMBLETON†
Physics Department, Imperial College, London

MS. received 16th *June* 1954 *and in amended form* 3rd *November* 1955

Abstract. Experimental results are discussed covering many aspects of the propagation and reflection of finite amplitude sound pulses and weak shock fronts. These are compared with theory and show amongst other features (*a*) that an N-shaped wave form tends to shorten on reflection and (*b*) that a shock front disappears on reflection at an open-ended tube.

---

§ 1. Introduction

The results of this paper represent an attempt to determine whether the theories of Stokes (1848), Earnshaw (1860), Rankine (1870), Rayleigh (1910), Taylor (1910) and others describe accurately the behaviour of sound pulses of finite amplitude. Pulses up to 5 cm of mercury pressure amplitude, travelling into undisturbed air at atmospheric pressure have been investigated experimentally. An apparatus suitable for this work has been described by Timbrell (1951); the present author has made a number of modifications with the purpose of simplifying the experimental technique and enabling results to be determined with greater accuracy. All measurements were made at a constant ambient air temperature of $20° \pm 1°\mathrm{C}$ and barometric pressure of 76 ± 1 cm of mercury. The sound pulse, which was a decaying wave train of sinusoidal form, was produced (Timbrell 1953) by the discharge of a condenser through the voice coil of a horn-type loudspeaker and was propagated as a one-dimensional wave (i.e. a wave defined by a single space coordinate x) through the air contained in a heavy, approximately circular, cylindrical steel tube of about 4 cm internal diameter. A beam of light was passed along a tube diameter through two plane optically perfect windows, which were flush with the smooth inside surface of this tube and which could be moved to any point along its length. This light beam represented one of the two interfering light beams of a Mach–Zehnder interferometer. As the sound pulse was propagated along the tube the air underwent small changes in density; when these traversed the light beam the fringe system formed by the interferometer moved by a proportional amount in a direction perpendicular to the length of the fringes. These fringes were arranged to be perpendicular to the axis of the propagation tube and to be formed as a virtual fringe system in the plane of this tube; by means of a lens a real image was thrown on to a narrow slit placed in front of a photomultiplier cell whose output was amplified and applied to the Y-plates of a cathode ray oscillograph. The vertical deflection of the cathode spot was a function of the density changes taking place in the sound pulse at the position of the light beam (more correctly the average density along the diameter of the shock tube at this point), this function being determined by the distribution of light intensity in the fringe system.

---

†Now at the Division of Applied Physics, National Research Council, Ottawa, Ontario, Canada.

The wave traces were recorded photographically, using a single stroke time base. To interpret these traces it was necessary to calibrate the horizontal deflection in terms of a known timing signal, and the vertical deflection in terms of known density, or pressure changes in the tube. The latter calibration was performed by using a semi-automatic electrical manometer (Embleton 1954) which enabled the vertical deflection of the cathode spot to be determined as a function of the excess pressure in the sound wave, above or below atmospheric pressure: this calibration was displayed on the oscillograph and recorded photographically. Hence it was possible to produce and measure a sound wave form as a function of time at any desired position along the axis of the propagation tube. By reproducing exactly the same pulse at the loudspeaker with the interferometer placed at some other point along the tube, any changes undergone by the pulse during its propagation may be studied.

All the wave forms discussed here have the general shape of one of the types shown in figure 1. In discussion of N-shaped waves, (figure 1(*c*)) it is usual for authors to refer to the two shock fronts in the wave as the 'head' and 'tail' shocks respectively and this system of nomenclature will be followed here. However, many authors also refer to that point in the pulse at which the excess pressure returns to zero at the end of the first half cycle as the tail of a triangular wave; to avoid confusion this point is here referred to as the 'origin'. The origin is

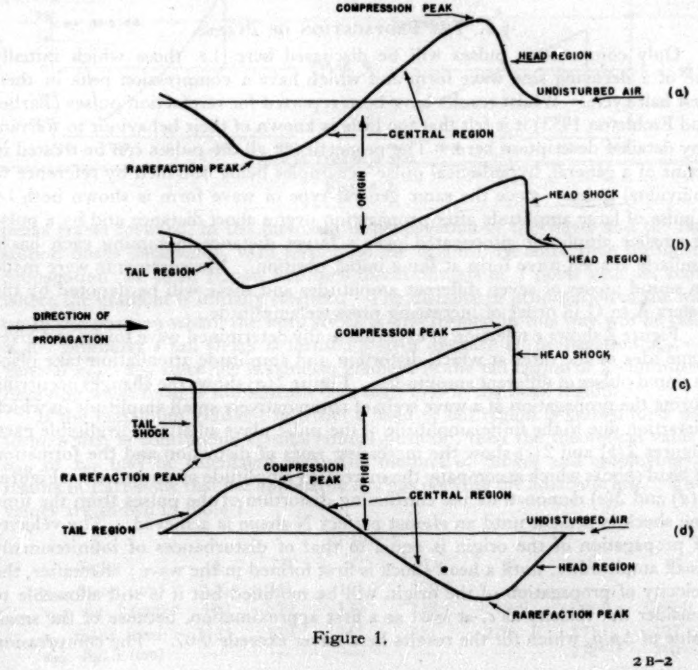

COMPRESSION PEAK
HEAD REGION
UNDISTURBED AIR
(a)
CENTRAL REGION
RAREFACTION PEAK
ORIGIN
HEAD SHOCK
(b)
HEAD REGION
TAIL REGION
DIRECTION OF PROPAGATION
COMPRESSION PEAK
HEAD SHOCK
(c)
TAIL SHOCK
RAREFACTION PEAK
COMPRESSION PEAK
ORIGIN
CENTRAL REGION
UNDISTURBED AIR
TAIL REGION
(d)
HEAD REGION
RAREFACTION PEAK

Figure 1.

taken as the zero of both the excess time Δt (the difference in time between the passage of any element of the wave and the passage of the origin past the point of measurement) and the excess pressure Δp coordinates in the pulse. The points of maximum excess pressure in the pulse are referred to as the 'compression' or 'rarefaction' peaks, when the excess pressure is positive or negative respectively. That part of the pulses discussed here can be divided into three 'regions'—ahead of the first peak is the 'head region'; between the peaks of the first and second half cycles and containing the origin is the 'central region', whilst between the peaks of the second and third half cycles is the 'tail region'. Upon distortion of a wave form of initial compression, as in figure 1 (*a*), the head and tail regions are those which ultimately contain the head and tail shocks respectively of an N-shaped wave form. All pulses are denoted by their total distances of propagation x, since genesis at the loudspeaker. When considering reflected pulses, the location of the reflector surface with respect to the loudspeaker is denoted by the distance r. It should be noted that the phenomena presented here relate to details of pulses of such small amplitude that they have no satisfactory representation as characteristic curves in the (x, t) plane which is conventionally used in the discussion of shock waves. For example the excess times Δt of the largest excess pressures discussed in this paper change at about 1% of the rate for the elapsed times t and it is with discrepancies in the rate of change in Δt, and quantities of similar magnitude that much of the present paper is concerned.

§ 2. THE PROPAGATION OF PULSES

Only compression pulses will be discussed here, i.e. those which initially are of a decaying sine wave form and which have a compression peak in their first half cycle. Whilst results have been reported for rarefaction pulses (Barber and Embleton 1953) it is felt that too little is known of their behaviour to warrant any detailed description here. The properties of all the pulses can be treated in terms of a general, hypothetical pulse—examples being obtained by reference to individual pulses—since the same general type of wave form is shown both by a pulse of large amplitude after propagation over a short distance and by a pulse of smaller amplitude propagated over a larger distance, assuming each has a similarly shaped wave form at some initial position. Measurements were made on sound pulses of seven different amplitudes and these will be denoted by the letters A to G in order of increasing pressure amplitude.

Figure 2 shows a selection of experimentally determined wave forms and gives some idea of the rates at which distortion and amplitude attenuation take place in sound pulses of different amplitudes. Figure 2 (*a*) shows the changes occurring during the propagation of a wave train of comparatively small amplitude in which distortion due to the finite amplitude of the pulse plays an almost negligible part. Figures 2 (*b*) and 2 (*c*) show the increasing rates of distortion and the formation of head shocks which accompany the increasing amplitude of the pulses. Figures 2 (*d*) and 2 (*e*) demonstrate the continuing distortion of the pulses from the time one shock is present until an almost perfect N-shape is achieved. The velocity of propagation of the origin is equal to that of disturbances of infinitesimally small amplitude c, until a head shock is first formed in the wave; thereafter, the velocity of propagation of the origin will be modified but it is still allowable to consider the velocity as c, at least as a first approximation, because of the small value of $\Delta p/p_0$ which for the results here never exceeds 0·07. The compression

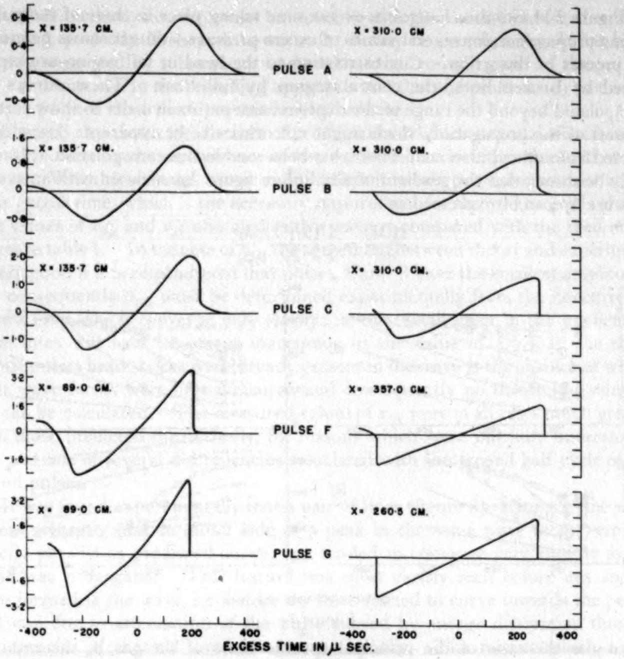


Figure 2.

peaks travel forward, in the direction of propagation of the wave, and the rarefaction peaks backwards, with respect to the origin. At some time during the propagation the head region becomes discontinuous at a point near its centre where the gradient is initially steepest. The distance of propagation of the wave up to the position where the head shock is first formed in this way will be called x_{SH}. Similarly a tail shock is formed after a distance of propagation x_{ST}. In general $x_{ST} > x_{SH}$ since the maximum gradient in the tail region of a continuous, decaying wave form is numerically less than that in the head region.

If the maximum gradient in either a head or tail region is known for a wave form which is continuous at some initial position, then the theoretical value of x_{SH} or x_{ST} may be calculated using the method of Stokes and employing the results of Earnshaw that are applicable to a gas obeying the adiabatic law. These values are shown in table 1.

Table 1

Pulse	A	B	C	D	E	F	G
x_{SH} theor. (cm)	526	241	131·7	120·6	—	—	—
x_{SH} fig. 3 (cm)	400	290	132	118	<89·0	<89·0	<89·0
x_{ST} theor. (cm)	1131	539	347	281	210	171	127
x_{ST} fig. 3 (cm)	—	—	—	—	—	270	210

Figure 3 shows the changes in excess time taking place in three of the pulses during propagation for several values of excess pressure—thus it shows primarily the process of distortion. Curves relating to the head or tail region are represented by broken lines, the central region by full lines. These curves are extrapolated beyond the range of the experimental points in order to show certain features of the propagation which might not otherwise be apparent: knowledge derived from other pulses, not shown, has been used in this extrapolation. It may easily be shown that the gradient of any line in figure 3 associated with an excess pressure Δp can be expressed as

$$\frac{\partial(\Delta t)}{\partial x} = \frac{1}{c}\,\frac{(\gamma+1)}{2\gamma}\,\frac{\Delta p}{p_0}$$

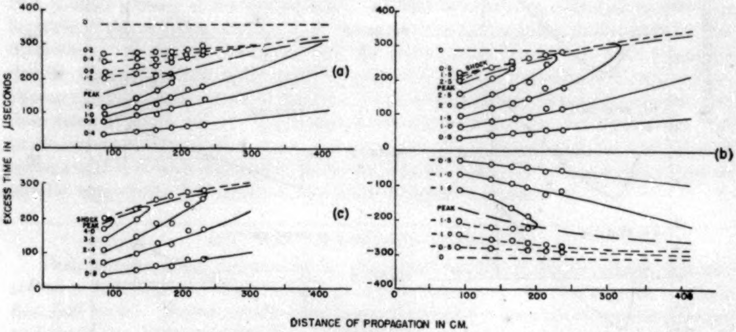


Figure 3.

where γ is the ratio of the principal specific heats of the gas, p_0 the ambient pressure and c the velocity of propagation of infinitesimally small disturbances. Hence the plotted curves, being drawn with Δp as parameter, should be straight lines. The gradient of each should be proportional to its value of Δp; and a pair of lines relating to the same Δp, one in say the head region and the other in the central region, should be parallel. Figure 3 shows that the experimental points lie approximately on straight lines provided that the excess pressures to which they refer are removed from the nearest peak by a certain fraction of the peak excess pressure. In the head region the fraction probably has the same value as in the central region up to the time a shock is formed but in this case an individual excess pressure then becomes absorbed into the head shock. It will so remain until the amplitude of the pulse has been attenuated to such an extent that this excess pressure becomes the upper excess pressure limit of the shock. As the pulse further attenuates this excess pressure leaves the shock and has a brief independent existence until it finally disappears at the compression peak; this represents the 'rounding-off' of the peaks which is evident in figure 2. The strongest head shocks measured in these experiments occupied 97–99% of the excess pressure change in the head region of their respective wave forms and there was nothing to suggest that shocks of up to 5 cm Hg pressure amplitude ever extended completely to the peak. As a pulse distorts the excess time of a peak at first approaches towards that of its corresponding shock but ultimately differs from it by a small and approximately constant amount. There appears to be no further tendency for the peaks to coincide with the shocks, i.e. for the pulse to become perfectly N-shaped, however far it may be propagated. It has been recently suggested by subsequent workers that the 'rounding-off' of the peak is due to a limitation in the recording apparatus and that the peaks are in fact more sharply pointed than these present results indicate.

Figure 3 also shows the distance of propagation at which a shock is first formed in a wave form by the concurrence of two or more lines in either the head or tail region. When this occurs two or more different excess pressures co-exist at the same excess time, which is the necessary requirement for the existence of a shock. The values of x_{SH} and x_{ST} obtained in this way are compared with the theoretical values in table 1. In the case of x_{SH} the agreement between theory and experiment is very good if it is remembered that pulses A and B have the smallest amplitudes and consequently x_{SH} must be determined experimentally from the concurrence of two lines which converge only slowly; a very small error in the gradient of these lines will lead to serious inaccuracy in the value of x_{SH}. In the three largest pulses head shocks were already present in the wave at the position at which their wave forms were first measured and consequently no theoretical value of x_{SH} can be calculated. The measured values of x_{ST} were in all cases much greater than those predicted theoretically, for reasons which were not fully understood; this was one of several discrepancies associated with the second half cycle of the sound pulses.

It was found experimentally that a pair of lines (figure 3) relating to the same excess pressure, one on either side of a peak in the same wave form were not exactly parallel as predicted above, but tended to converge very slightly as the pulse was propagated. This feature was most clearly seen before any shocks were formed in the wave, i.e. before the lines started to curve towards the peaks, and was due to attenuation of the pulse caused by energy dissipation through the agencies of thermal conduction and viscosity. The rate at which a pair of lines converge may be related to the amplitude attenuation coefficient of the pulse. In principle it is therefore possible to compare the attenuation coefficients derived from the various excess pressure levels with those from the peaks within the same pulse—and also to compare the attenuation coefficient derived from a certain excess pressure when at a peak with that when removed from the neighbourhood of a peak. In practice the gradients of the lines could not be determined with sufficient accuracy to allow this comparison to be made accurately, though it was found that, before a shock was formed in the pulse, all were of the same order of magnitude; for the first half cycle the experimental results varied from 5×10^{-4} to $12 \times 10^{-4}\,\text{cm}^{-1}$, in the second half cycle the results were somewhat higher and varied from 11×10^{-4} to $36 \times 10^{-4}\,\text{cm}^{-1}$.

The amplitude attenuation of the peaks, both before and after a shock is present, may be determined from the envelopes of the curves in figure 4 which, if drawn for the same pulses would present the same information as figure 3 but in a different form; these values of attenuation coefficient are shown in table 2. Figure 4, obtained by plotting the changes in excess pressure during propagation at a series of excess times shows primarily the amplitude attenuation within the pulse but the presence of distortion and shock formation is also evident. In the central region distortion results in an increased 'attenuation' of the various curves which otherwise would have a smaller gradient than the corresponding peak line at the same distance of propagation. In the head and tail regions the

Table 2. Amplitude Attenuation Coefficients α

Pulse	A	B	C	D	E	F	G
Compression peak × 10^4	9·1	12	7·8	7·8	—	—	—
	—	—	22·7	30·7	24·7	27·4	33
Rarefaction peaks × 10^4	10·2	18	21·9	14·5	3·9	10·6	15

Pulse A

Δp	0·1	0·2	0·3	0·4	0·5	0·6	0·7	−0·1	−0·2	−0·3
$\alpha \times 10^4$	11·4	6·9	6·6	7·8	7·8	9·7	10·3	36	21	18

Pulse B

Δp	0·2	0·4	0·6	0·8	1·0	—0·2	—0·4	—0·6
$\alpha \times 10^4$	—	—	—	5·3	10·3	18·1	11	11

effects of distortion outweigh the amplitude attenuation which exists as a result of energy dissipation and at excess times lying in these regions the excess pressure shows an increase. When a shock, which increases its separation from the origin during propagation, passes through the value of excess time associated with one of the curves in figure 4 then the excess pressure coordinate of this curve shows an abrupt change.

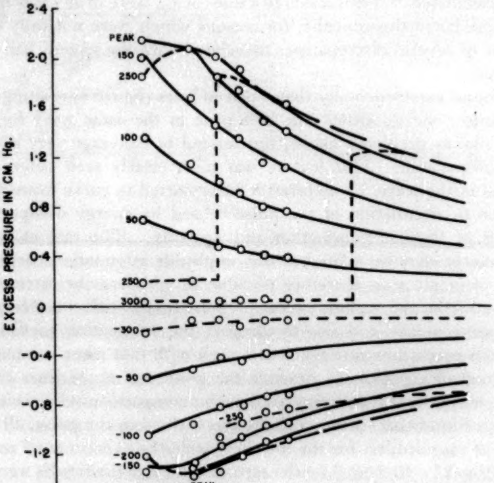


Figure 4.

The difference in gradient of a pair of lines (figure 3) gives rise to a spurious difference in the particle velocity, calculated from these gradients and associated with this excess pressure. Considering the first half cycle of the wave the particle velocity appears increased above its true value in the central region and decreased in the head region. The particle velocity, in terms of the gradient of a line in figure 3 is given by

$$u = \frac{2c^2}{(\gamma+1)} \frac{\partial(\Delta t)}{\partial x}.$$

Similarly the acoustic impedance of the medium is

$$\frac{\Delta p}{u} = \frac{(\gamma+1)}{2c^2}\,\frac{\Delta p}{\partial(\Delta t)/\partial x}$$

which also has different values when calculated from different regions of the wave form. The values of $\partial(\Delta t)/\partial x$ are obtained directly as the gradient of lines in figure 3, for which the corresponding Δp's are the parameters of these lines. Some numerical values of particle velocity and acoustic impedance are given in table 3. Values are given only for those regions of a wave form which are apparently unaffected by the presence of shocks in other parts of the pulse, i.e. where there is no curving of the excess pressure lines in figure 3 towards their respective peak lines. Amplitude attenuation of the pulse should lead to an increase in the apparent acoustic impedance measured from a head or tail region and to a decrease if measured from a central region. Few results are available from the head regions of the measured wave forms because of the formation of shocks but the impedance calculated from the head region in pulse A is considerably greater than that calculated from the central region and also greater than the theoretical value of 42 c.g.s. units. From inspection of figure 2(*a*) the negative impedance calculated from the tail region of pulse A may easily be recognized as a large increase in the positive impedance caused by attenuation of the wave form. Similarly the various values of impedance calculated from the central regions of the wave forms are in general less than the theoretical value of 42 c.g.s. units. If allowance is made for the modification of the results caused by amplitude attenuation of the wave form, the experimentally determined values of acoustic impedance are in agreement with the theoretical value. However, for this same reason it is not possible to use this method for the determination of impedance since amplitude attenuation is inherent in all wave propagation

Table 3. Experimental Values of Excess Pressure, Particle Velocity and Acoustic Impedance

Pulse	Δp (cm Hg)	Head region		Central region		Tail region	
		Particle vel. (cm sec^{-1})	Impedance (c.g.s. units)	Particle vel. (cm sec^{-1})	Impedance (c.g.s. units)	Particle vel. (cm sec^{-1})	Impedance (c.g.s. units)
A	0·6	112	70·5	290	27·2	—	—
	0·4	94	55·6	160	33	—	—
	0·2	28	94	84	31·2	—	—
	−0·2	—	—	−84	31·2	197	−133
B	1·0	—	—	553	23·8	—	—
	0·8	356	29·5	412	25·5	—	—
	0·4	216	24·4	216	24·4	—	—
	−0·4	—	—	−178	29·5	—	—
C	1·0	—	—	422	31·2	—	—
	0·5	—	—	178	37	—	—
	−0·5	—	—	−150	44	0	∞
D	1·5	—	—	562	35	—	—
	1·0	—	—	338	39	—	—
	0·5	—	—	188	35	—	—
	−0·5	—	—	−169	39	−56·2	117
	−1·0	—	—	−375	35	−234	56
E	2·4	—	—	937	34	—	—
	1·6	—	—	600	35	—	—
	0·8	—	—	328	32	—	—
	−0·8	—	—	−244	43	—	—

(energy dissipation) and is an essential feature of the propagation of finite amplitude pulses which contain shock waves (due to lengthening of the pulse).

The excess velocity of a head shock, determined from the slope of the shock line in figure 3, was considerably greater than the value predicted theoretically from Rankine's relation

$$c\frac{(\gamma+1)}{4\gamma}\frac{\Delta p}{p_0}.$$

This was so if the peak excess pressure was employed for the purposes of calculation and the discrepancy was even greater if only the change in excess pressure in the abrupt part of the head region was taken. However, if the peak has been rounded off by some limitation of the recording apparatus and is in fact more angular and of greater amplitude, then the discrepancy becomes much smaller. Furthermore it has been assumed that the velocity of the origin is at all times *c*, but this is strictly true only up to a time just later than that at which the head shock is first formed. This latter point may also explain yet a further discrepancy in the behaviour of the second half cycle of the wave forms—once a tail shock has been formed it does not increase its separation from the origin as rapidly as predicted from Rankine's theory. This is seen very clearly in figure 5 which relates to pulse F after an approximately N-shaped wave form has been obtained: the head shock is rapidly increasing its separation from the origin but the tail shock shows only a very slight increase.

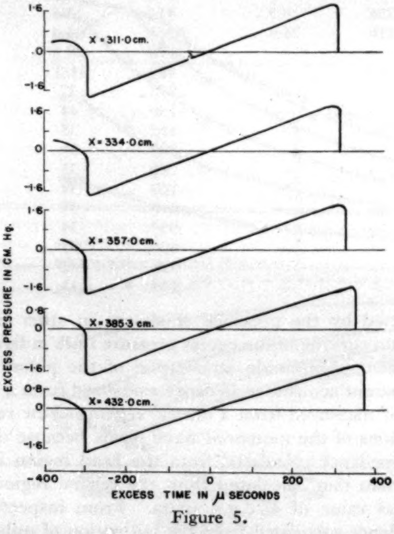


Figure 5.

When considering the propagation of a wave form of the type discussed here, it is not allowable to consider each region separately. This is achieved in Friedrichs' (1948) theory which takes into account the interaction of one region with another. It is probable that the presence of shocks within the wave will modify the propagation in the immediate neighbourhood of the origin. On this assumption the point referred to in this paper as the origin will not be propagated with the normal sonic velocity *c* after a certain length of time has elapsed from the formation of a shock in the wave form. In this event the value given here for the excess velocity of a head shock will be modified; if it were reduced, assuming Rankine's theoretical expression is correct, the excess velocity of the corresponding tail shock would be altered in such a way that its experimentally determined velocity would also approach more nearly its theoretical value.

§ 3. Reflection of Pulses

3.1. *In a Tube with a Closed End*

The solid material normally used for reflecting the sound pulses was brass, in the form of a cylindrical plug 10 cm long. In order that sound pulses might be propagated over different distances before reflection several cylindrical extension tubes were used, of the same internal diameter as the propagation tube, into the ends of which could be inserted the reflecting plug. This plug fitted each tube closely and when packed round with vaseline, to make the system airtight, it required a steady force of 5 to 10 lb wt to draw it from its tube. Experiments were also carried out with reflecting plugs of varying lengths and composed of other materials; plugs of copper, teak, balsa wood, sponge and magnesium oxide were used covering a range of thermal diffusivity from 10^{-1} to $3 \times 10^{-4}\,\mathrm{cm^2\,sec^{-1}}$. With the exception of those reflected from the sponge, all the wave forms were identical within the limits of experimental error. It therefore appears that the thermal diffusivity of a reflector surface has a negligible influence on the reflected wave and that the important criterion is that the surface should be free from pores of any appreciable size. It was also found that the properties of the reflected wave were unaffected by varying the length of the reflector plug between 2·5 and 10 cm, but changes would probably occur if the plug were so thin that it behaved as a membrane and did not remain plane under the impulse of an incident shock.

The incident and reflected wave forms of each sound pulse were measured at a series of positions along the propagation tube, at a sufficiently great distance from it that each reflected wave form was obtained without the superposition of its corresponding incident wave (and vice versa). The reflected wave forms measured by the interferometer could not be determined with as great an accuracy as the incident wave forms. Even when observations were made 1 metre from the reflector the reflected wave arrived at the interferometer before the ambient pressure in the tube had returned to a steady state. The recorded wave forms could largely be corrected for this disturbance but nevertheless there was some uncertainty (of an amount up to about 20 μsec) in locating the exact point in the pulse to be taken as origin. The recorded wave forms were used to plot graphs of the types shown in figures 3 and 4; these were then extrapolated to the position of the reflector surface. In this way the incident and reflected wave forms were obtained as they would exist at the reflector, each without the superposition of the other.

Figure 6 (*a*) and (*b*) shows typical pairs of corresponding incident and reflected wave forms and it can be seen that the general type of the wave form is unchanged by reflection at a solid surface. Thus the wave form is reflected without change in the sign of the excess pressure at a boundary of increased acoustic impedance, as is to be expected from theoretical considerations. Also, if a shock is present in any region of an incident wave form, then a shock is also present in the same region of the corresponding reflected pulse. There is nothing in the experimental results to indicate that the reflection coefficient was influenced to any measurable degree by the amount of distortion present in the wave form. Inspection of corresponding pairs of incident and reflected wave forms showed that (1) the amplitude reflection coefficient, i.e. the ratio of reflected to incident pressure amplitude from peak to peak, is in general greater than unity and (2) there is a tendency for the wave forms to shorten on reflection, i.e. for the length of the wave in terms of excess time, from peak to peak, or from head to tail shock, to become less. These two effects are shown in figure 7 for one position of the reflecting plug but are typical of results obtained at all positions of the plug. The plotted points show a large scatter which may be attributed to the large number of graphical processes to which the experimental results have been subjected in order to arrive at these points. However, the phenomenon of shortening itself cannot be attributed to experimental error since in a small number of cases the reflected wave forms at the point of measurement were shorter than the corresponding incident wave forms at the same point by an amount several

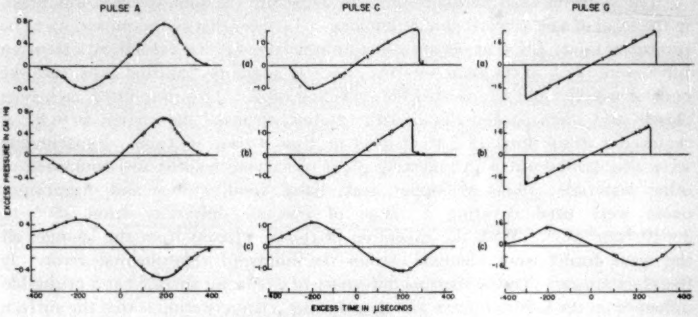
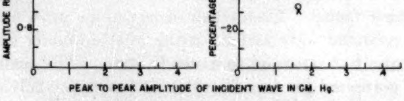


Figure 6.



(a) (b)

Figure 7.

times greater than the estimated experimental error in measurement of excess time. In figure 7 (*b*) the percentage change in length of a wave is shown for both the separation between the head and tail shock (or midway along the head and tail regions if a shock is not present) and also between the compression peak and the rarefaction peak. Of 40 such points, 4 indicated an increase in length, 3 indicated no change whilst 33 showed a decrease in length. The points show too great a scatter to allow any correlation to be attempted between the amplitude of the wave and its percentage shortening.

If a wave form on reflection retains its same general shape then a shortening of the wave form appears to be a necessary consequence of any increase in amplitude during the reflection process since there is no source of energy present at the reflector. An analysis by Finkelstein (1947) indicates that for a simple triangular wave form propagated in a gas the pressure on a solid reflector surface is prolonged in duration, the shape of the wave form is altered and the reflected wave form is increased in length, which is contrary to the experimental results discussed here. An explanation of the process of shortening of the wave form upon reflection has been sought (Embleton 1952) by an analysis similar in form to that of Finkelstein but considering the process in greater detail. It shows that during part of the process of reflection a shortening of the wave does in fact occur although this is nullified by an increase in length at a later stage. The total lengthening predicted by this theory is much less than that predicted by Finkelstein, but does not explain the shortening of the wave form which is found experimentally. The theory indicates that any sound pulse, having a first half cycle of triangular form, should lengthen by a fraction of its length equal to

$$\frac{u^2(3-\gamma)(\gamma+1)}{2\{4c+(\gamma-1)u\}\{4c-(3-\gamma)u\}}$$

where u is the particle velocity associated with the peak of the wave form. For wave forms of the amplitudes investigated here this fraction is negligibly small (of the order of 0·01%).

3.2. *In a Tube with an Open End*

For the purpose of obtaining wave forms propagated back along a tube after reflection at an open end only the propagation tube itself was used, without any extension tube. The open end of the tube was flush with, and surrounded by, a heavy concentric cylinder of steel, about 13 cm in diameter. For the purposes of the present description it may therefore be assumed that the open end is effectively a circular orifice in an infinite plane wall at least for the first half cycle of the pulse.

The wave forms as they existed at the open end (in the absence of the incident wave) were obtained by extrapolation, as were the waves reflected from a closed end. Figure 6 (*a*) and (*c*) shows typical pairs of corresponding incident and reflected waves and it may be seen that at an open end the wave forms are reflected with the sign of the excess pressure reversed in each half cycle. This is to be expected on theoretical grounds since such a phase change is necessary between the incident and reflected wave in order to satisfy the boundary condition of an open end that the excess pressure should at all times be zero.

Chester (1950) has recently discussed the reflection of a sound pulse at the open end of a two-dimensional channel, bounded by two parallel planes, in which it is being propagated. In principle his analysis may be applied to the one-dimensional case of reflection at the open end of a tube. He shows that the reflected wave may be expressed as the sum of a series of singly and multiply reflected and diffracted waves from the edges of the open end and the walls of the tube. None of these wave fronts is perpendicular to the axis of the tube (except at an infinite distance from the open end) but continually cross each other as they are reflected in the walls of the tube. From this it is not surprising that no evidence of any shock exists in the reflected wave forms. Even if the original diffracted waves, which are the only source of the reflected waves (formed when the incident wave form reaches the open end of the tube), were discontinuous their energy is so dissipated by propagation according to the inverse square law and generation of secondary diffracted waves that they soon lose their discontinuous character. If such a discontinuity takes the form of a rarefaction shock it is an unstable system and immediately becomes continuous.

It is reasonable to assume that the reflected wave forms were approximately plane and perpendicular to the axis of the propagation tube by the time they were recorded, since they were propagated back from the open end a distance of about twelve times the diameter of the tube before the first measurements of them were made. Thus the measured wave forms were at least a close approximation to the reflected waves at the point of measurement, although it is doubtful, according to Chester's analysis of the problem, whether the wave forms at the open end obtained by extrapolation bear any resemblance to a wave system which has a physical existence. There is a tendency for the reflected wave forms in figure 6 not to recross the line of zero excess pressure in the neighbourhood of $-400\,\mu$sec of excess time: this is probably due to the 'unallowable' extrapolation of measured wave forms back to the open end of the tube; any unusual features would be most likely to occur in the tail region since this is formed by the sum of several wave systems whereas the head region may have only one or two component systems.

§ 4. CONCLUSION

The results described in this paper allow a reasonably consistent understanding of the properties of finite amplitude sound pulses to be obtained. This is in qualitative agreement with current theory but quantitatively there are discrepancies on several points. The amplitude attenuation coefficient associated with the second half cycle is less than that associated with the first half cycle in each pulse. Also the experimentally determined shortening of a wave form upon reflection at a closed end is contrary to any change in length predicted by current theory.

It is desired to emphasize that in finite amplitude pulses the phenomena of particle velocity, distortion of the wave form, shock formation, amplitude attenuation and acoustic impedance are all interrelated: none of these may be considered in any detail without reference to the others.

ACKNOWLEDGMENTS

The author wishes to express his thanks to Dr. R. W. B. Stephens for his great interest, encouragement and criticism throughout the work and to Professor Sir George Thomson for the use of the facilities of the Physics Department of Imperial College. He is indebted to the Ministry of Education and the

University of London for grants which enabled the work to be carried out; also to the Dixon Research Fund of London University for the purchase of a Southern Instruments recording camera.

REFERENCES

BARBER, G. J., and EMBLETON, T. F. W., 1953, *Nature, Lond.*, **172**, 1057.

CHESTER, W., 1950, *Phil. Mag.*, **41**, 11.

EARNSHAW, S., 1860, *Phil. Trans. Roy. Soc.* A, **150**, 133.

EMBLETON, T. F. W., 1952, *8th Int. Congr. Appl. Mech., Istanbul;* 1954, *Rev. Sci. Instrum.*, **25**, 246.

FINKELSTEIN, R., 1947, *Phys. Rev.*, **71**, 42.

FRIEDRICHS, K. O., 1948, *Commun. Appl. Math.*, **1**, 211.

RANKINE, W. J. M., 1870, *Phil. Trans. Roy. Soc.*, **160**, 277.

RAYLEIGH, LORD, 1910, *Proc. Roy. Soc.* A, **84**, 247.

STOKES, G. G., 1848, *Phil. Mag.*, **33**, 349.

TAYLOR, G. I., 1910, *Proc. Roy. Soc.* A, **84**, 371.

TIMBRELL, V., 1951, *Nature, Lond.*, **167**, 306 ; 1953, *Ibid.*, **172**, 540.

Luminescence and Electrical Conductivity associated with Defects in Alkali Halide Crystals

By J. EWLES and J. C. STEAD†
Department of Physics, The University of Leeds

MS. received 9*th August* 1955 *and in amended form* 10*th October* 1955

Abstract. The emission and excitation spectra of the luminescence produced in potassium and sodium chlorides by thermal shock, mechanical strain or by the introduction of the divalent ions Ba, Sr or Ca show similarities in the same host crystal which are explicable on the assumption that the positive ion vacancies comprise or are part of the luminescent centre. Differences in the emission spectra in the same host crystal activated in these various ways are attributed to differences of association of the vacancies. The electrical conductivity of pure annealed material as determined by an a.c. method is increased by thermal shock or mechanical strain. The similarity of the gradients of the graphs showing log conductivity plotted against reciprocal temperature of the thermally or mechanically strained samples with those of the annealed samples containing added divalent ions shows that the increased conductivity is due to positive ion vacancies; the magnitude of the increase as compared with the pure annealed crystals indicates that the number of 'free' vacancies is increased some four or five times by the thermal shock.

§ 1. Introduction

There is now evidence that lattice defects may determine the efficiency and even the appearance of luminescence in crystals (Ewles and Lee 1953). It is also generally accepted that in many cases one type of defect, lattice vacancies, is responsible for the electrical conductivity of ionic solids. Such vacancies are increased in number with rise in temperature and may be 'frozen in' by quenching from a high temperature. They may also be increased in number by the introduction of an ion different in valency from those of the host crystal. The object of the experiments described in this paper was to investigate any correlation between the luminescent and electrical properties of some monovalent ionic solids in which defects had been deliberately introduced by thermal treatment or by the introduction of heterovalent ions. The materials investigated were 'Specpure' potassium chloride and sodium chloride and the heterovalent ions introduced as desired were barium, strontium or calcium.

The luminescent emission and excitation spectra and the electrical conductivity over a range of temperature were examined for (i) pure annealed single crystals, (ii) pure material quenched suddenly *in vacuo* from the melting point to the temperature of liquid nitrogen, (iii) pure material sublimed *in vacuo* into a platinum crucible cooled to the temperature of liquid nitrogen and (iv) annealed single crystals from melts containing small additions ($\sim 6 \times 10^{-4}$ atomic concentration) of barium, strontium or calcium.

---

† Now at the Admiralty Research Laboratory, Teddington, Middx.

§ 2. EXPERIMENTAL ARRANGEMENTS

2.1. *Materials*

All materials were of 'Specpure' quality. Single crystals were grown from the tip of a platinum thimble immersed in a melt contained in a platinum crucible which was very slowly cooled (about 1°c min^{-1}).

2.2. *Luminescence*

The *emission spectra* were excited by a 125 watt high-pressure Mercra mercury vapour lamp with the outer glass envelope removed and optically filtered to eliminate all wave lengths between 3341 Å and 5779 Å by the Bowen liquid filter (Bowen 1935). The spectra were recorded on Kodak OB or IF plates by means of a Hilger E2 spectrograph and photometered on a recording photometer to be described elsewhere.

The *excitation spectra* were investigated by a modification of the now well-known method of photoelectric fluorescent photometry previously described by one of us (Ewles 1949) using a hydrogen arc as source of continuous ultra-violet exciting light, a Hilger D 246 monochromator for dispersion and an E.M.I. 5060 photomultiplier together with a comparison target of constant quantum efficiency (sodium salicylate) as receiver, feeding signals into a d.c. electrometer amplifier. Scattering in the monochromator and dark current in the photomultiplier were allowed for by the interposition of a glass plate for each zero reading and scattering of the exciting light by the target was eliminated by the interposition of an appropriate glass filter between this target and the photomultiplier.

2.3. *Electrical Conductivity*

In any d.c. method for determining electrical conductivity of ionic solids, measurements are complicated by polarization effects depending on the potential applied and the time for which it is applied. In previous work a.c. methods have been limited by stray capacitance effects to resistances of not more than about 10^7 ohm so that investigations could only be carried out with alkali halides either at temperatures above about 300°c or with the unsatisfactory d.c. method. It was found possible, however, to construct a Schering bridge with further modification of the ingenious arrangement of Hall (1952) which, by the elimination of stray capacitances, enabled resistances up to at least 10^{11} ohms to be measured with an a.c. method (Dawson and Stead 1954). The specimen, in the form of either a single crystal or a compressed tablet of the order of 1 cm^2 area and 1 mm thick was held between platinum-rhodium discs serving as electrodes. Platinum–platinum-rhodium thermocouple wires welded to the back of these discs passed through capillaries in flat ended silica rods. This assembly was contained in a silica tube considerably wider in the middle than the specimen and narrowed down at the ends so as to fit snugly the silica rods holding the specimen. Springs between hooks fused to the silica rods and the silica tube held the assembly and crystal specimen firmly in place. The whole was contained in an electric furnace.

§ 3. THE EXPERIMENTAL RESULTS

3.1. *Excitation Peaks*

The excitation peaks are shown graphically in figure 1. In all cases the treatment of the crystals whether thermal, mechanical or by addition of divalent ions produces an extension of the fundamental excitation tail of the pure annealed crystals to at least 2200 Å.

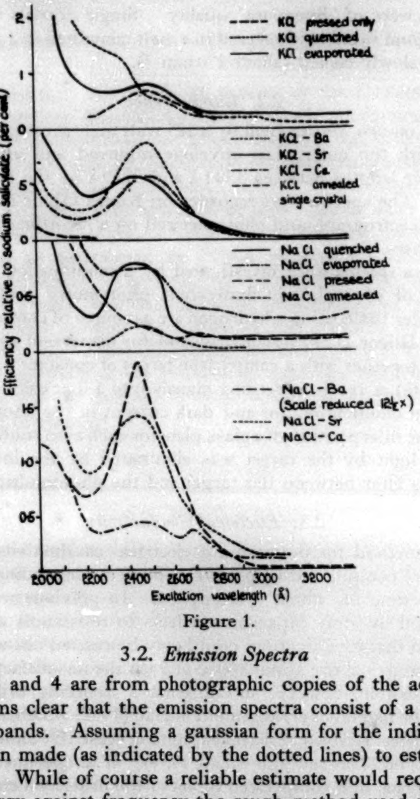


Figure 1.

3.2. *Emission Spectra*

Figures 2, 3 and 4 are from photographic copies of the actual photometer curves. It seems clear that the emission spectra consist of a number of overlapping broad bands. Assuming a gaussian form for the individual bands, an attempt has been made (as indicated by the dotted lines) to estimate the course of these bands. While of course a reliable estimate would require a replotting in terms of energy against frequency the rough method used should give some idea of the nature and number of the component bands.

3.3. *Conductivity*

Observations have been confined in this work to sodium chloride. The nature of the results are indicated in figures 5, 6, 7 and are discussed in the next section. More detailed work on potassium chloride, pure and containing various concentrations of barium, has been carried out in this laboratory and will be reported later.

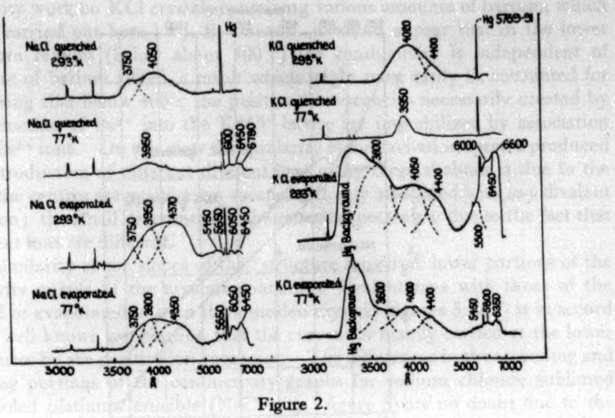


Figure 2.

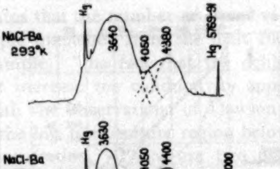


Figure 3.

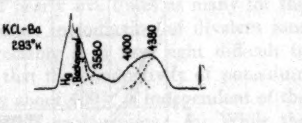


Figure 4.

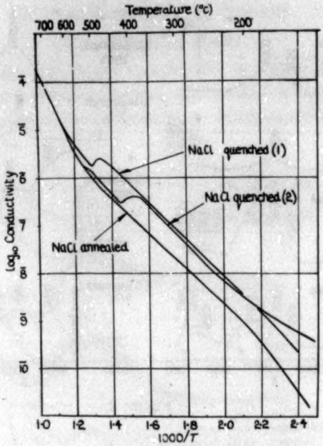
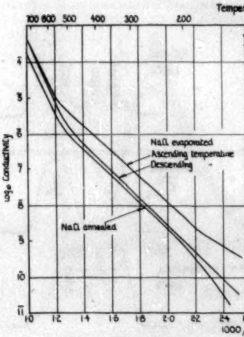


Figure 6.



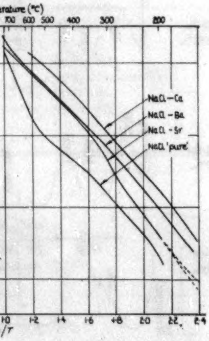


Figure 5. Figure 7.

§4. DISCUSSION OF RESULTS

The close similarity of the excitation peaks of KCl pure quenched, KCl pressed, KCl annealed single crystals containing Ca or Sr or Ba and of the excitation peaks of NaCl pure quenched, annealed NaCl single crystals containing Ca or Sr or Ba, suggests that in each case the excitation occurs in a common centre. The fact that the addition of a divalent ion to a monovalent lattice increases the efficiency of this common excitation and is well known to increase the number of positive ion vacancies strongly suggests that the common centre is the positive ion vacancy or a complex associated with it.

In the emission spectra, although there are similarities, e.g. between KCl : Sr and KCl : Ba at 77°K, the differences are too marked to be ignored. In the conductivity work on KCl crystals containing various amounts of barium, which has been carried out here by J. B. Dawson, it would appear that in the lower temperature regions (below about 400°c) the conductivity is independent of the amount of barium added, a result which might most easily be accounted for by supposing that below 400°c the positive ion vacancies necessarily created by the introduction of Ba^{2+} into the K^+Cl^- lattice are immobilized by association with the Ba^{2+} ions. On this view the similarity of the excitation spectra produced by the introduction of different divalent ions or by thermal shock is due to the fact that the centres are positive ion vacancies closely associated with *any* divalent positive ion; the small differences in the *emission* spectra are due to the fact that the divalent ions are different.

The similarity of the slopes of the 'structure sensitive' lower portions of the conductivity curves of the crystals containing divalent ions with those of the quenched or evaporated or even the annealed crystals (figures 5, 6, 7) is in accord with the well-known supposition that the current is mainly carried at the lower temperatures by the positive ion vacancies. The differences in the ascending and descending portions of the conductivity graphs for sodium chloride sublimed into a cooled platinum crucible (NaCl evap., figure 5) are no doubt due to the solution in the lattice of the vacancies produced by thermal shock by the annealing process of slow heating and cooling during the course of the measurements. A comparison of the conductivities with that of the 'pure' annealed crystal at say 300°c indicates that the number of 'free' vacancies is nearly four times as many for the sample quenched from the melt and nearly five times as many for the sublimed sample. The fact that the deliberate introduction of divalent ions does in fact increase the conductivity appreciably is at first sight difficult to reconcile with the observations of Dawson that the conductivity of potassium chloride in the low temperature region below about 400°c is independent of the barium concentration. There are two possible explanations: (*a*) While the extra positive ion vacancies necessarily created by the introduction of a divalent ion are prevented by association from taking part in conduction in this temperature region, the existence of the dipoles (divalent positive ion and positive ion vacancy) disturbs the surrounding lattice and so makes it easier for thermal vacancies to be created or, in other words, lowers the energy barrier for the migration of normal lattice ions. It is interesting to note in this connection that mechanical and thermal strains undoubtedly produce a luminescence attributable to positive ion vacancies and that such strain induced luminescence is very greatly enhanced if the substance contains heterovalent ions. (A fuller report on these effects is in preparation.) (*b*) The divalent ion has a limiting solubility in the lattice at the lower temperatures, separating out as $BaCl_2$ for instance.

It is possible that both (*a*) and (*b*) are true.

§ 5. CONCLUSIONS

It is considered that this work affords support for the ideas that (i) thermal or mechanical strains produce or facilitate the production of lattice vacancies, (ii) positive ion vacancies either alone or in association with lattice ions or with the foreign ions or with both can act as luminescent centres, (iii) the nature of the foreign ion, if this is part of such a centre, only slightly affects the excitation and emission energy levels involved in the excitation and emission of luminescence.

ACKNOWLEDGMENTS

We are grateful to the Department of Physics of the University of Leeds for facilities afforded and to the Government Grant Committee of The Royal Society for a grant with which some of the apparatus used was purchased. One of us (J.C.S.) is indebted to the University of Leeds for a maintenance grant.

REFERENCES

BOWEN, E. J., 1935, *J. Chem. Soc.*, 35.
DAWSON, J. B., and STEAD, J. C., 1954, *Proc. Leeds Phil. Soc.*, **6,** 133.
EWLES, J., 1949, *Proc. Leeds Phil. Soc.*, **5,** 236.
EWLES, J., and LEE, N., 1953, *J. Electrochem. Soc.*, **100,** 392.
HALL, H. C., 1952, *J. Sci. Instrum.*, **29,** 224.

RESEARCH NOTES

Low Temperature Magnetoresistance Anomalies in Indium Antimonide

By I. M. MACKINTOSH
Department of Physics, University of Nottingham

Communicated by L. F. Bates ; MS. received 2nd November 1955

Recent investigations (Hung and Gliessman 1950, Fritzsche 1955) have indicated that at temperatures below about 10°K germanium exhibits anomalous behaviour of the Hall constant R, the resistivity ρ, and the magnetoresistive ratio $\Delta\rho/\rho_H$, ρ_H being the resistivity in an applied magnetic field H. These investigations of the electrical properties of semiconductors at low temperatures have been extended to indium antimonide (Fritzsche and Lark-Horovitz 1955) and essentially similar behaviours of the Hall constant and resistivity have been observed. A point of particular interest, however, is that with the latter material negative values of the magnetoresistive ratio have been observed at temperatures below about 8°K, for which no explanations have yet been offered.

The anomalies in germanium have been attributed by Hung (1950) to a small but finite conduction in the impurity band at low temperatures, and this model can account qualitatively for the observed behaviour of germanium, although a detailed theory of impurity band conduction has not so far been evolved. It is doubtful whether Hung's model alone could predict a negative magnetoresistive ratio, and in this note an attempt is made to show that a mechanism which might be important at low temperatures, and which might explain the phenomenon, is provided by the changes in the concentrations of charge carriers which occur when the impurity levels supplying these carriers are split in a magnetic field.

The analysis assumes a theoretical model of a semiconductor with a set of localized impurity (donor) levels at an energy E below the bottom of the conduction band E_c. The application of a magnetic field H splits these impurity levels into two sets of localized levels displaced by an energy ΔE from their original position. Neglecting any possible conduction in an impurity band, the conductivity in the absence of a magnetic field will be given by

$$\sigma = n\boldsymbol{e}\mu \qquad \ldots\ldots(1)$$

where $\boldsymbol{e}$ is the electronic charge, n the density of conduction electrons and μ their mobility. The change in the conductivity when a field is applied will then be due to changes in both n and μ, i.e.

$$\Delta\sigma = \Delta\mu\frac{\partial\sigma}{\partial\mu} + \Delta n\frac{\partial\sigma}{\partial n} \qquad \ldots\ldots(2)$$

which gives

$$\frac{\Delta\rho}{\rho_H} = -\frac{\Delta\sigma}{\sigma} = -\left(\frac{\Delta\mu}{\mu} + \frac{\Delta n}{n}\right). \qquad \ldots\ldots(3)$$

Thus $\Delta\rho/\rho_H$ may take negative values if the relative increase in the population of conduction electrons $\Delta n/n$ exceeds the relative decrease in their mobility $\Delta\mu/\mu$.

If N is the density of impurity levels, and also the density of electrons in these levels at 0°K, E_F is the Fermi level, and m_n the effective mass of electrons in the conduction band, then for Fermi–Dirac statistics the situation at any temperature T may be represented (Hutner *et al.* 1950) by

$$\frac{2N_c}{\sqrt{\pi}}F(\eta)+\frac{N}{1+\exp\{(E-E_F)/\boldsymbol{k}T\}}=N \qquad \ldots\ldots(4)$$

where

$$N_c=2(2\pi m_n\boldsymbol{k}T/\boldsymbol{h}^2)^{3/2}; \qquad \eta=E_F/\boldsymbol{k}T$$

and

$$F(\eta)=\int_0^\infty \frac{x^{1/2}}{\exp(x-\eta)+1}\,dx. \qquad \ldots\ldots(5)$$

Values of $F(\eta)$ between -4 and $+20$ are tabulated by McDougall and Stoner (1929). In equation (4) the first term represents the density of electrons in the conduction band and the second term is the density of electrons in the impurity levels.

The distribution of electrons in the presence of a magnetic field is similarly given by

$$\frac{2N}{\sqrt{\pi}}F(\eta)+\frac{\frac{1}{2}N}{1+\exp\{(E+\Delta E-E_F)/\boldsymbol{k}T\}}+\frac{\frac{1}{2}N}{1+\exp\{(E-\Delta E-E_F)/\boldsymbol{k}T\}}=N. \qquad \ldots\ldots(6)$$

The position of the Fermi level can then be found in each case by plotting the left-hand sides of equations (4) and (6) as functions of E_F and equating them to N. The density of conduction electrons is subsequently found by substituting these values of E_F in the term $(2N_c/\sqrt{\pi})F(\eta)$. An alternative method of analysis yielding essentially the same results, is given by Stevens (1956).

In order to make a qualitative comparison with the results obtained by Fritzsche and Lark-Horovitz with their purest material (specimen *c*), the values $N=10^{16}\,\mathrm{cm}^{-3}$ and $E=0{\cdot}5\times10^{-3}$ ev were taken. The analysis was then carried out at various temperatures between 2°K and 10°K assuming three different values for ΔE. It should be noted, however, that the experiments were actually carried out on p-type indium antimonide, but this should not detract greatly from the value of such a comparison.

The magnitude of the splitting to be expected is calculated from the expression

$$\Delta E=\tfrac{1}{2}g\beta H \qquad \ldots\ldots(7)$$

where β is the Bohr magneton, and g is the Landé splitting factor, assumed to have the value 2. The values of ΔE used in the calculations were $0{\cdot}025\times10^{-3}$, $0{\cdot}05\times10^{-3}$ and $0{\cdot}10\times10^{-3}$ ev corresponding approximately to fields of 4000, 8000 and 16000 oersteds respectively.

The normal magnetoresistance effect in cubic semiconductors was investigated theoretically by Seitz (1950). From his general expression, the transverse magnetoresistive ratio is found to be

$$\frac{\Delta\rho}{\rho_H}=\frac{9\pi}{16\mathbf{c}^2}\left(1-\frac{\pi}{4}\right)\mu^2H^2. \qquad \ldots\ldots(8)$$

Assuming that this expression is valid at low temperatures, and neglecting any population changes due to the magnetic field, from equations (3) and (8) we have

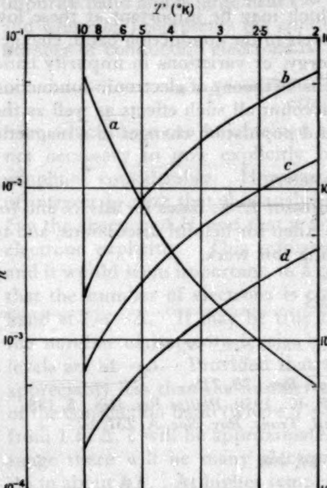
$$\frac{\Delta\mu}{\mu}=-\frac{\Delta\rho}{\rho_H}=-3{\cdot}8\times10^{-17}\mu^2H^2. \qquad \ldots\ldots(9)$$

Experimental data on the mobility in indium antimonide are not available for temperatures much below 10°K. For the purposes of comparison it was therefore assumed that $\mu = 10^3$ cm volt^{-1} sec^{-1} at 10°K, in close agreement with the value obtained for specimen *c* by Fritzsche and Lark-Horovitz, and that at lower temperatures ionized impurity scattering predominates and the mobility takes the form

$$\mu = AT^{3/2} \qquad \ldots\ldots(10)$$

where A is a constant.

Figure 1 shows the assumed temperature variation of $-\Delta\mu/\mu$ and the calculated curves of $\Delta n/n$ for the three values of ΔE. The region of curve *d* at temperatures above about 3°K is somewhat uncertain since these very small changes in $\Delta n/n$ represent almost the limit of accuracy of the graphical method used to determine Δn. The theoretical curve of $(\log \Delta\rho/\rho_H, 1/T)$ for $\Delta E = 0{\cdot}025 \times 10^{-3}$ ev as calculated from equation (3), and the experimental curve for specimen *c*, are shown in figure 2.



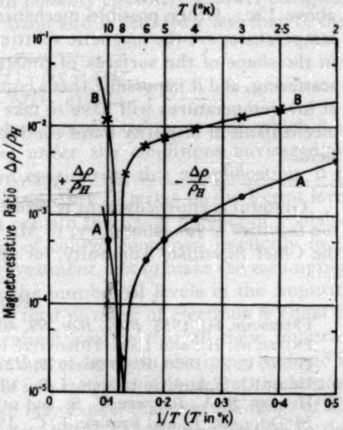


Figure 2. Variation of log $\Delta\rho/\rho_H$ with $1/T$. A, theoretical; B, experimental (Fritzsche and Lark-Horovitz 1955, specimen *c*).

Figure 1. Variation of $\log(-\Delta\mu/\mu)$ with $1/T$ (curve *a*), and $\log \Delta n/n$ with $1/T$ for $\Delta E = 0{\cdot}10 \times 10^{-3}$, $0{\cdot}05 \times 10^{-3}$, and $0{\cdot}025 \times 10^{-3}$ ev (curves *b*, *c* and *d* respectively).

The magnitude of the relative population changes in a simple model of this type are seen to be quite small. The normal magnetoresistance effect is also very small at these low temperatures, however, and there is a qualitative agreement between the theoretical curve of figure 2 ($H = 4000$ oersteds) and the experimental curve for specimen *c*, obtained by Fritzsche and Lark-Horovitz with $H = 3500$ oersteds.

The following points arise in considering the results of these calculations:

(1) There is a discrepancy between the shapes of the two curves and between their magnitudes at any particular temperature.

(2) The magnitude of the calculated negative magnetoresistive ratio increases approximately as the square of the magnetic field at any particular temperature, whereas Fritzsche and Lark-Horovitz have observed a linear dependence for fields greater than about 1500 oersteds.

(3) The effect of a higher density of impurities, neglecting any effect this may have on the width of the impurity band, will cause greater relative changes in population at any given temperature, i.e. greater values of $\Delta n/n$, and will therefore tend to move the zero of the magnetoresistance curves to higher temperatures. This is observed experimentally.

The discrepancies in magnitude and shape of the curves and in the character of the field-dependence of the effect are serious evidence against the theory of magnetic splitting of the impurity levels as an explanation for the negative magnetoresistance effect. Nevertheless it is thought to be of interest to report the perhaps surprising result that quite small changes in magnetic energies may result in a reversal of sign of the magnetoresistive ratio at temperatures above 1°K. Other possible mechanisms which may be important at these low temperatures are the magnetic splitting of levels in the conduction band, changes in the shape of the surfaces of constant energy, or variations in impurity band scattering, and it is possible that a comprehensive theory of electronic conduction at low temperatures will have to take into account all such effects as well as the mechanisms of impurity band conduction and population changes in a magnetic field.

ACKNOWLEDGMENTS

Grateful acknowledgment is made to Professor L. F. Bates for advice and for the facilities of his laboratory, to Mr. J. W. Allen for helpful discussions, and to the Chief Scientist, Admiralty, for sponsoring this work.

REFERENCES

FRITZSCHE, H., 1955, *Phys. Rev.*, **99**, 406.
FRITZSCHE, H., and LARK HOROVITZ, K., 1955, *Phys. Rev.*, **99**, 400.
HUNG, C. S., 1950, *Phys. Rev.*, **79**, 727.
HUNG, C. S., and GLIESSMAN, J. R., 1950, *Phys. Rev.*, **79**, 726.
HUTNER, R. A., RITTNER, E. S., and DU PRE, F. K., 1950, *Philips Res. Rep.*, **5**, 188.
McDOUGALL, J., and STONER, E. C., 1929, *Phil. Trans. Roy. Soc.* A, **237**, 67.
SEITZ, F., 1950, *Phys. Rev.*, **79**, 372.
STEVENS, K. W. H., 1956, *Proc. Phys. Soc.* B, **69**, 406.

Population Changes with Magnetic Field

BY K. W. H. STEVENS
Department of Physics, University of Nottingham

Communicated by L. F. Bates ; MS. received 2nd November 1955

I HAVE seen the preceding note (Mackintosh 1956) prior to publication and it seems of interest to present an alternative approximate method of finding $\Delta n/n$. In place of Mackintosh's equation (6) we consider instead the more general relation:

$$N = \tfrac{1}{2} \int \left[\frac{f(\epsilon + \beta H) + f(\epsilon - \beta H)}{\exp\{-(E_F - \epsilon)/\boldsymbol{k}T\} + 1} \right] d\epsilon$$

where $f(\epsilon)$ represents the density of all the levels, conduction and impurity, in the absence of the magnetic field. This equation, which defines E_F, is unaltered on reversing the sign of H, so we may expect to be able to expand E_F in a power series in H^2, thus:

$$E_F = \zeta + bH^2 + \ldots$$

Assuming that $bH^2 \ll \boldsymbol{k}T$ it is then found that

$$b = -\tfrac{1}{2}\boldsymbol{k}T\beta^2 \int \left[\frac{d^2f/d\epsilon^2}{\exp\{-(\zeta-\epsilon)/\boldsymbol{k}T\}+1}\right] d\epsilon \Bigg/ \int \left[\frac{f(\epsilon)}{\exp\{-(\zeta-\epsilon)/\boldsymbol{k}T\}+1}\right] d\epsilon.$$

The numerator can be expressed alternatively, by integration by parts, as

$$\frac{-\beta^2}{2\boldsymbol{k}T} \int \frac{f(\epsilon)\exp\{-(\zeta-\epsilon)/\boldsymbol{k}T\}[1-\exp\{-(\zeta-\epsilon)/\boldsymbol{k}T\}]}{[\exp\{-(\zeta-\epsilon)/\boldsymbol{k}T\}+1]^3} d\epsilon$$

showing that the condition $bH^2 \ll \boldsymbol{k}T$ is equivalent to $\beta H \ll \boldsymbol{k}T$. If we split $f(\epsilon)$ into two parts, $g(\epsilon)$ and $h(\epsilon)$, where $g(\epsilon)$ represents the density of levels in the impurity band and is large near $\epsilon = -\Delta$ and possibly elsewhere at lower energies, and where $h(\epsilon)$ represents the conduction band and is zero for ϵ less than 0, the number of conduction electrons is

$$n + \Delta n = \tfrac{1}{2} \int \frac{h(\epsilon+\beta H) + h(\epsilon-\beta H)}{\exp\{-(\zeta + bH^2 - \epsilon)/\boldsymbol{k}T\}+1} d\epsilon$$

from which $\Delta n/n$ can be obtained as a rather complicated expression which it is not necessary to give explicitly because, under the conditions envisaged, it simplifies considerably. However, before considering this simplification, it is of interest to note that this analysis expresses $\Delta n/n$ in terms of the Fermi level ζ in the absence of H, and that it does not contain the actual number of available electrons explicitly. This information is of course contained implicitly in ζ, and it would seem important, in a general treatment, not to make the assumption that the number of electrons is equal to the number of levels in the impurity band at $\epsilon = -\Delta$. It may be true that the total number of electrons is equal to the number of impurity levels, but we cannot assume that all these impurity levels are at $-\Delta$. Provided that the number of levels in the impurity band is appreciably less than the number of levels in a range of order $\boldsymbol{k}T$ at the bottom of the conduction band (where $T \sim 1$°K) we may expect that over the range of $\boldsymbol{k}T$ from 1 to Δ, ζ will be approximately equal to $\boldsymbol{k}T$. This follows because in this range there will be many electrons in the conduction band occupying levels up to about $\boldsymbol{k}T$. At higher temperatures ζ will not rise so rapidly, and below it will eventually fall below zero. In the range of interest, with $\zeta = \boldsymbol{k}T$, the expression for $\Delta n/n$ is found to approximate to $-A\beta^2H^2/\boldsymbol{k}^2T^2$, where A varies only slowly with temperature and is of order unity. The width of the impurity band proves to be relatively unimportant provided that it is much less than Δ (not βH as might perhaps be expected) and we may sum up by observing that it is Δ which controls the temperature at which $\Delta n/n$ begins to grow from zero and that its magnitude is determined by $\beta^2H^2/\boldsymbol{k}^2T^2$. For temperatures above 1°K this is quite small, but so also is $\Delta\mu/\mu$, and as Mackintosh has shown in the preceding note it is possible to get a reversal in the sign of the magnetoresistance, though the observed variation with H suggests that some other mechanism is also operative.

REFERENCE

MACKINTOSH, I. M., 1956, *Proc. Phys. Soc.* B, **69**, 403.

The Influence of ^{3}He on the Lambda Point

By BIPIN KUMAR AGARWAL
Physics Department, Allahabad University, Allahabad, India

MS. received 18*th October* 1955, *and in final form* 23*rd November* 1955

THE presence of ^{3}He in a mixture of ^{3}He and ^{4}He shifts the lambda point T_λ of the mixture towards temperatures lower than $T_\lambda{}^0$, that for pure ^{4}He. This was shown by Abraham, Weinstock and Osborne (1949) for mixtures with moderate ^{3}He concentrations of 2% to 25%, and by Daunt and Heer (1950) for mixtures with high concentrations of 40% to 90%. Similar measurements were also made by Esellson and Lazarev (1950) for concentrations less than 2% and recently by King and Fairbank (1953) for concentrations less than 4%.

Several theoretical attempts have been made to calculate the dependence of the lambda point on the ^{3}He concentration. In the theories of de Boer and Gorter (1950), and of Rice (1949, 1950), Engel and Rice (1950), a thermodynamic evaluation is made of the change of the lambda point with ^{3}He concentration, on the basis of the Taconis hypothesis (Taconis *et al.* 1949). Daunt and Heer (1951) suggested that ^{3}He mixtures can be regarded as due to an ideal mixture of a degenerate Bose–Einstein (BE) gas and a non-degenerate Fermi–Dirac (FD) gas. Assuming statistical independence of the BE and FD gases in solutions of ^{3}He in ^{4}He and a perfect gas spectrum $E=p^2/2m$, they showed that the variation of the lambda temperature can be predicted in rough agreement with the observed data. The perfect gas spectrum is presumably distorted in the liquid state. The hypothesis is accordingly made that the spectrum is determined by

$$E=Ap^{1/r},$$

where $A=1/2m$ and r is a simple function of the concentration $X=N_3/(N_3+N_4)$ of ^{3}He atoms, N_3 and N_4 denoting the number densities of ^{3}He and ^{4}He atoms in the mixture, respectively. In accordance with the calculations of Daunt and Heer it follows then that

$$\frac{T_\lambda}{T_\lambda{}^0}=\left[\frac{N_4V_4}{N_3V_3+N_4V_4}\right]^{1/3r} \qquad \ldots\ldots(1)$$

or in terms of ^{3}He concentration X,

$$\frac{T_\lambda}{T_\lambda{}^0}=\left[\frac{1-X}{1+X(V_3/V_4-1)}\right]^{1/3r} \qquad \ldots\ldots(2)$$

where T_λ is the lambda temperature of mixture, $T_\lambda{}^0$ is the lambda temperature of pure ^{4}He, and V_3 and V_4 are the known molar volumes of ^{3}He and ^{4}He, respectively. The dependence of r on concentration X is taken as follows,

$$r=r_0\,\frac{N_3V_3}{N_3V_3+N_4V_4} \quad \text{for} \quad X>0{\cdot}5, \qquad \ldots\ldots(3)$$

and

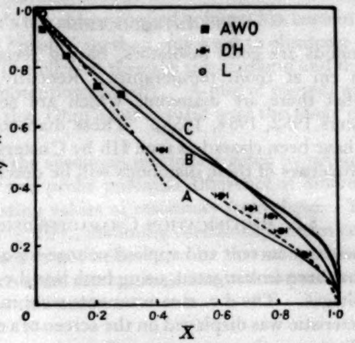
$$r=r_0\,\frac{N_4V_4}{N_3V_3+N_4V_4} \quad \text{for} \quad X<0{\cdot}5, \qquad \ldots\ldots(4)$$

where $r_0=0{\cdot}5$, the value for the perfect gas spectrum. A comparison with experiments and with other existing theoretical evaluations is given in the figure.

The slope of the $(T_\lambda/T_\lambda{}^0, X)$ curve for very low concentrations of ^{3}He can be written as

$$\frac{\partial T_\lambda}{\partial X} = -\frac{T_\lambda{}^0}{r}\frac{V_3}{V_4}. \qquad \ldots\ldots(5)$$

For a concentration of 2·4% the value of r given by equation (4) is 0·484, which when substituted in equation (5) gives a shift of −2·04 degrees per mole. This may be compared with the recent measurements of King and Fairbank (1953) who found that for concentrations below 4% this shift is of −1·5 degrees per mole.



The ratio of T_λ, the lambda temperature of a mixture to $T_\lambda{}^0$ that of pure ^{4}He as a function of concentration X.
AWO, experimental results of Abraham, Weinstock and Osborne. DH, experimental results of Daunt and Heer (1950), L, Leiden specific heat results.
Curve A, de Boer and Gorter (1950), curve B, de Boer (1949), curve C, ideal mixture, $r = 0·5$. The broken curve shows the present results.

The more recent measurements of Dash and Taylor (1955) also yield a slope of the same order for low concentrations. For a concentration of 20·3% the value of r is 0·371 and that of the shift is −2·67 degrees per mole which is the same as given by de Boer and Gorter for lower concentrations of this order. The general agreement of the two computed branches of the curve with the observed data is also good.

Further applications of this model have been worked out and will be published elsewhere.

REFERENCES

ABRAHAM, B. M., WEINSTOCK, B., and OSBORNE, D. V., 1949, *Phys. Rev.*, **76,** 864.
DE BOER, J., 1949, *Phys. Rev.*, **76,** 852.
DASH, J. G., and TAYLOR, R. D., 1955, *Phys. Rev.*, **99,** 598.
DAUNT, J. G., and HEER, C. V., 1950, *Phys. Rev.*, **79,** 46 ; 1951, *Ibid.*, **81,** 447.
ENGEL, O. G., and RICE, O. K., 1950, *Phys. Rev.*, **78,** 55.
ESELLSON, B. N., and LAZAREV, B. G., 1950, *Dokl. Akad. Nauk, S.S.S.R.*, **72,** 265.
GORTER, C. J., and DE BOER, J., 1950, *Phys. Rev.*, **77,** 569.
KING, J. C., and FAIRBANK, H. A., 1953, *Bull. Amer. Phys. Soc.* (3), **28,** 65.
RICE, O. K., 1949, *Phys. Rev.*, **76,** 1701 ; 1950, *Ibid.*, **77,** 142 ; 1950, *Ibid.*, **79,** 1024.
RICE, O. K., and ENGEL, O. G., 1950, *Phys. Rev.*, **78,** 183.
TACONIS, K. W., BEENAKKER, J. J. M., NIER, A. O., and ALDRICH, L. T., 1949, *Phys. Rev.*, **75,** 1966.

Electrical Measurements on Type IIb Diamonds

By H. B. DYER and P. T. WEDEPOHL
Department of Physics, University of Reading

MS. received 14*th September* 1955

§ 1. Introduction

Most diamonds are good insulators, having resistivities of the order of 10^{16} ohm cm at room temperature. Recently, however, it has been shown that there are diamonds which are semiconductors at room temperature (Custers 1952, 1954, 1955). These diamonds are usually, but not always, blue, and have been classed as type IIb by Custers. Work done on some of the electrical properties of these diamonds will be described in this paper.

§ 2. Rectification Characteristics

The relation between current and applied voltage for a metal–semiconducting diamond contact has been investigated, using both low d.c. voltages and comparatively high a.c. voltages. The d.c. measurements were made potentiometrically, and the a.c. characteristic was displayed on the screen of a cathode-ray oscilloscope (see figures 1 and 2).

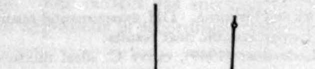


Figure 1. Low voltage d.c. rectification characteristic,

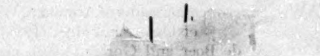


Figure 2. High voltage a.c. rectification characteristic.

The non-rectifying base contact was a layer of evaporated silver, and the metal used for the point contact a thin tungsten wire. Only at certain points on the diamond surface was the probe contact resistance low enough for measurements to be made.

The sign of rectification suggests that the conduction is predominantly p-type. The rectification ratio at 1 volt is approximately 60, and there is a breakdown of the rectification at 60 to 70 volts.

§ 3. VALIDITY OF OHM'S LAW

The validity of Ohm's law at low applied voltages has been investigated for two specimens. The usual 'probe' method was used to eliminate the effect of non-ohmic base contacts. The current through the diamond and the voltage between probes were measured potentiometrically. A plot of current against voltage was strictly linear, showing that Ohm's law is valid over the range of voltages applied (0–4 volts).

The homogeneity of the specimen was investigated by taking readings at room temperature of current and probe potential difference at different (known) probe separations, and calculating values of resistivity from these. The resistivity was found to be 80 ohm cm ± 10%, assuming constant specimen cross section. The range of measured values is not believed to be due to specimen inhomogeneity, as variations in specimen cross section, together with experimental error, result in an uncertainty of about this order.

§ 4. TEMPERATURE DEPENDENCE OF CONDUCTIVITY

The bulk resistance of one specimen (D 106) has been measured at temperatures ranging from 200°K to 1040°K. Again a potentiometric 'probe' method was used, and the applied voltages were within the range for which Ohm's law had been proved to be valid.

A plot of log r against $1/T$ (r = bulk resistance, T = absolute temperature) is linear in the range 200°K to 290°K, and shows that the resistance has a minimum at about 590°K (see figure 3). The slope of the linear portion of the curve gives a dependence of $2\boldsymbol{k}\ln r$ upon $1/T$ of 0·77 ev ($\boldsymbol{k}$ = Boltzmann's constant). This value is related to the activation energy but should not be assumed to be equal to the activation energy.

§ 5. HALL COEFFICIENT

Preliminary measurements of the Hall coefficient at room temperature showed the specimen (D 106) to be a p-type semiconductor. The carrier concentration at room temperature was found to be 8×10^{13} carriers cm^{-3}, assuming only p-type conduction to be present.

§ 6. CORRELATION WITH OPTICAL DATA

In all, eight specimens were examined and a correlation found between the conductivity and the intensity of the blue colour. It has been shown (Clark, Ditchburn and Dyer 1956) that the blue colour is caused by one of a series of bands extending into the infra-red, observed only in type IIb diamonds. Accurate measurements of these bands in four specimens confirms that the conductivity at room temperature is related to the infra-red band strength.

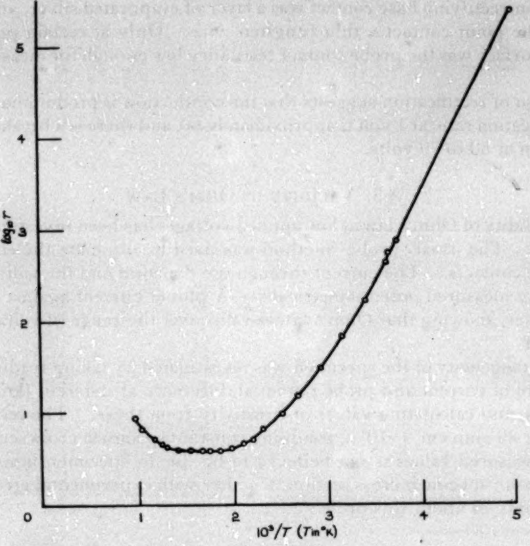


Figure 3. Resistance of type IIb diamond as a function of the inverse absolute temperature.

ACKNOWLEDGMENT

We would like to thank Dr. J. F. H. Custers for supplying the diamonds used in the investigation and for discussion of the relation between his own experiments and the present work.

REFERENCES

CLARK, C. D., DITCHBURN, R. W., and DYER, H. B., 1956, *Proc. Roy. Soc.* A, in the press.

CUSTERS, J. F. H., 1952, *Physica*, **18,** 489; 1954, *Ibid.*, **20,** 183; 1955, *Nature, Lond.*, **176,** 173.

REVIEWS OF BOOKS

Optical Glassworking, by F. TWYMAN. Pp. viii + 275. (London: Hilger & Watts Ltd., 1955.) 24*s*.

One cannot, I believe, make an optical surface with written instructions alone. In many books the detailed instructions are omitted—there is much which must be learnt by trial and error. Although not intended for the amateur, Twyman's book *Optical Glassworking* does in fact give some excellent detailed descriptions of procedure. The book is an abridged version of the second edition of the author's *Prism and Lens Making*, and has been written by request to 'meet the needs of those engaged in, or studying glass-working'. The account of 'Single surface working' gives descriptions of cutting, edging, grinding and polishing a prism and a singlet lens, including many of the wrinkles which one has at one's finger tips but often finds difficult to, or forgets to, record.

The author has drawn freely from other published work and some may feel that frequent reference to optical workshop practice of the early part of this century makes the book somewhat out of date. This, I think, is not so, for many features of the art of optical glassworking have changed little over the years. Indeed the two chapters on 'Testing optical work' could well have included a more lengthy description of the Foucault knife edge test.

Several writers have contributed to the book. There is an appendix by S. J. Underhill, on 'Making polarising prisms', a chapter on the 'Production of spectacle lenses' by A. A. S. Moore, and one on 'Microscope lenses' by R. J. Bracey, the latter containing information on pinhole testing which would help users of microscopes.

The book is intended primarily for optical technicians, and it is a pity that an abstract from a lecture is all that is included in a chapter of some four pages in length devoted to non-spherical surfaces. It may be that such work was considered out of place in this book, but on the other hand, there is a chapter of some thirty pages on 'Large object glasses' taken from the work of Draper, Grubb and Ritchey, with an account by J. V. Thompson of the 200-inch Mount Palomar mirror.

There are chapters on the production of prisms and lenses in quantity and one on the deposition of reflecting and anti-reflecting films.

This book will be of great assistance to optical technicians, and some will be glad to find that throughout the text names and addresses of suppliers of materials and machines are included, and that there is a list of reference books on optical glassworking. W. J. BATES.

Ultrasonic Engineering (*with particular reference to high power applications*), by A. E. CRAWFORD. Pp. x + 344. (London: Butterworths Scientific Publications, 1955.) 45*s*.

This book describes methods of generating and applications of ultrasonic waves, from the engineer's point of view. The first chapter summarizes the main features of ultrasonic propagation and the second discusses cavitation. The next four chapters describe piezoelectric, magnetostrictive, and electromagnetic generators, and whistles and sirens as ultrasonic sources. In a further six chapters there are described numerous applications of high power ultrasonics for such purposes as precipitation of smokes and fogs, emulsification of immiscible liquids, soldering of metals and inactivation of enzymes and in chemical processing. The last chapter deals with ultrasonic instruments such as flaw detectors and echo sounders.

The treatment of most topics is brief and descriptive but references are given at the end of each chapter for further reading. A great deal of the information is new and could only be found elsewhere, if at all, by diligent reading of original papers and patent specifications. For the most part information of practical use to the engineer is given in preference to theory. One would conclude that the author decided not to overlap previous books on ultrasonic physics, and this seems to have left the book a little unbalanced. Considering the many topics included it is surprising to find no mention of the increasing applications of ultrasonic delay lines in radar and electronic computers.

The author writes clearly, and the text is well illustrated by photographs and diagrams; there is an index. The book can be recommended to engineers and physicists who want to find out how to generate ultrasonic waves of high intensity and what they can be used for. J. M. M. P.

Radio Astronomy, by J. L. PAWSEY and R. N. BRACEWELL. Pp. x + 361. (Oxford: Clarendon Press, 1955.) (International Monographs on Radio.) 55*s*.

By 1945 only a few observations had been made that could be described as radio astronomical, yet today there are at least twenty receiving stations in operation all over the world, there have been a great many papers published, and the subject is regarded as so important by astronomers that in August a conference was specially organized at Jodrell Bank in connection with the triennial congress of the International Astronomical Union. There is in addition a good deal of general interest in the new branch of science and the time is certainly ripe for the appearance of this book, the first comprehensive text on the subject.

Dr. Pawsey directs the radio astronomy section of the Radiophysics Laboratory of C.S.I.R.O. at Sydney, where much of the research has been carried out, and Dr. Bracewell has also been working at Sydney so the authors are well qualified to present the balanced account that is needed in such a volume. They " have tried to make it readable by anyone with a background of physics " and by including some of the radio fundamentals for astronomers and introductory chapters on solar physics and astrophysics have, on the whole, succeeded in their aim.

The main topics considered are techniques of observation, theory of radio waves in ionized gases, solar radio waves, and cosmic radio waves, with shorter chapters on thermal radio waves from the moon, extraterrestrial radio echoes, meteors and atmospheric effects, and they are very competently dealt with, but it seems a pity that with one or two minor exceptions the subject-matter is only taken to the middle of 1952. It should surely be possible to get a manuscript into print with less than a three-year delay, this being particularly important when knowledge is accumulating as rapidly as it is in radio astronomy at present. If the book were written this year for example, one of the largest sections would have to be concerned with the observations of the neutral hydrogen line at 1421 Mc/s. In the present volume only nine pages appear on this work. The chapter on cosmic radio waves and discrete sources would also need to be much longer.

There are some half a dozen slight errors but only on page 131 could a reader be seriously misled, by some wrong numerical values. A surprising omission is any reference to what now seems one of the most plausible mechanisms of production of the radio waves in certain cases, namely, the acceleration of relativistic electrons in magnetic fields.

The book then is clear, readable and well-produced (as would be expected from the O.U.P.) and very welcome as an introduction to radio astronomy, but it is to be hoped that another edition is in preparation to include the great amount of work published since 1952. J. R. SHAKESHAFT.

Sonics, by T. F. HUETER and R. H. BOLT. Pp. xi + 456. (London: Chapman and Hall; New York: John Wiley, 1955.) 80*s*.

Optics, mechanics, electronics and now sonics! This latter word has been coined by the authors to describe " the technology of sound as applied to problems of measurement, control and processing ". It is a little unfortunate that the term ' ultrasonics ' has come to imply both the pure and technological aspects of high frequency acoustics.

In their most delightful book the authors, both leading workers in the acoustic world, commence with a brief introduction; this is followed by two chapters devoted to the development of the basic theory necessary as the background for the later chapters which deal with transducers and the various physical aspects and techniques of sonic processing. A noteworthy balance between theory and practice is maintained throughout the book; the general pattern is exemplified by the sections on piezoelectricity and magnetostriction, which start with a summary of the essential theory, followed by a discussion of the factors governing its application to practice and end with a description of relevant experimental work and a comprehensive list of references.

The authors have avoided a common pitfall of cataloguing indiscriminately the numerous chronicled ultrasonic effects and applications, spurious and otherwise, and are to be congratulated on the wisdom of their choice which contains examples representative of the latest developments in the subject.

A useful appendix to the book gives a very clear account of acoustical relaxation mechanisms in fluids and describes the use of acoustical techniques in yielding information about fundamental physical processes. Many books have been written on acoustics in recent years but it can be truly said that *Sonics* makes a definite contribution to the literature of the subject and it should be a treasured possession both of the student and of the advanced worker in sound.

R. W. B. S.

CONTENTS OF SECTION A

PAGE

Dr. E. **Corinaldesi**. The Two-body Problem in the Theory of the Quantized Gravitational Field 189

Dr. K. K. **Damodaran** and Mr. R. M. **Curr**. Single Scattering of Fast Electrons by Atomic Nuclei 196

Mr. L. J. **Tassie**. Nuclear Scattering of High Energy Electrons on the Independent-particle Model 205

Mr. R. **Batchelor**. Experiments on the Inelastic Scattering of Neutrons by Uranium 214

Dr. F. A. **El Bedewi**. Magnetic Analysis of the Inelastically Scattered Deuterons and Disintegration Protons from the Deuteron Bombardment of Fluorine . 221

Dr. E. H. **Kerner**. The Band Structure of Mixed Linear Lattices 234

Mr. J. N. **Murrell** and Dr. J. A. **Pople**. The Intensities of the Symmetry-forbidden Electronic Bands of Benzene 245

Dr. R. W. **Nicholls** and Mr. W. R. **Jarmain**. *r*-Centroids : Average Internuclear Separations Associated with Molecular Bands 253

Mr. D. L. **Booth**, Dr. G. **Preston** and Dr. P. F. D. **Shaw**. The Cross Section and Angular Distributions of the D–D Reactions between 40 and 90 kev . . 265

Dr. E. H. S. **Burhop** and Mr. R. **Marriott**. The Destruction of Metastable Helium Atoms by Collision-Induced Radiation 271

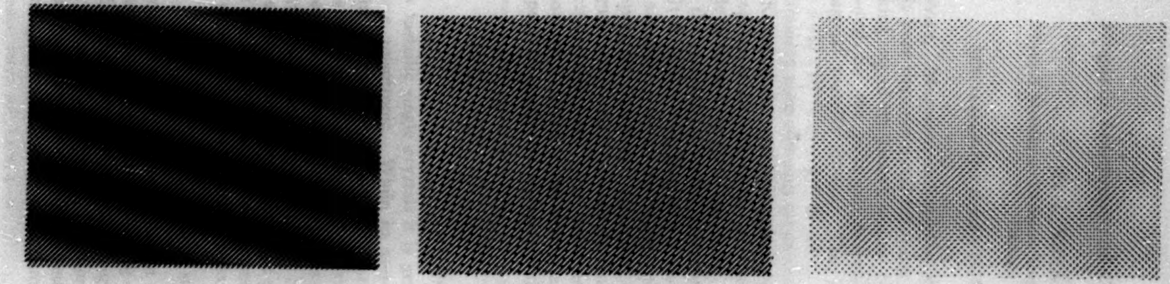
Research Note :

Mr. D. J. **Thouless**. Energy Levels of π-mesic Atoms 280

Letter to the Editor :

Dr. A. H. **Cooke** and Mr. J. G. **Park**. Nuclear Spins and Magnetic Moments of ^{161}Dy, ^{163}Dy, ^{171}Yb and ^{173}Yb 282

Contents of Section B 284



(*a*) (*b*) (*c*)

Figure 1. Examples of moiré (much enlarged).

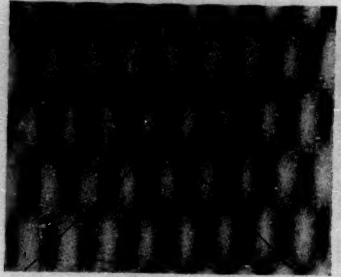


Figure 6. Example of homogeneous strain.

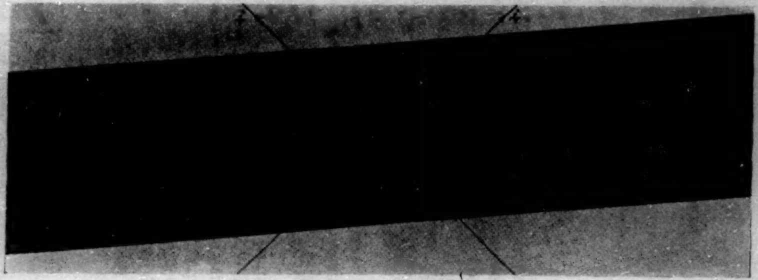


Figure 7. Example of heterogeneous strain.

Magnetic Anisotropy and Cold Worked Texture of Titanium

BY J. REEKIE AND Y. L. YAO
Physical Metallurgy Division, Department of Mines and Technical Surveys,
Ottawa, Ontario, Canada

MS. received 20*th October* 1955, *and in revised form* 5*th December* 1955

Abstract. Magnetic susceptibility determinations have been made for commercially pure titanium metal following various mechanical and thermal treatments. Making use of the known texture of heavily rolled titanium, values have been deduced for the principal susceptibilities and magnetic anisotropy of the hexagonal metallic lattice. The values obtained are: $\chi_{\parallel}=3{\cdot}34_0\times10^{-6}$ and $\chi_{\perp}=3{\cdot}08_7\times10^{-6}$ e.m.u. g^{-1} at 25°c. These figures have been confirmed by preliminary measurements on an oriented single crystal of titanium.

Knowing the magnetic anisotropy it is possible to estimate the texture in a specimen of the cold worked or annealed metal from susceptibility measurements made in three orthogonal directions. This has been done in a number of cases, and the practical use of this aspect of the work is discussed.

It is shown from magnetic measurements that heavily compressed titanium has a texture in which the basal planes of the hexagonal lattice are normal to the compression direction. Assuming that a compressed specimen approximates magnetically to single crystal behaviour, the temperature variations of the principal susceptibilities have been determined from 25°c to 300°c. The magnetic anisotropy is found to be temperature independent over this range, and the temperature coefficient of susceptibility, $d\chi/dT$, is $0{\cdot}0011\times10^{-6}$ e.m.u. g^{-1} deg^{-1}.

§ 1. INTRODUCTION

In metallic single crystals it is generally true that many physical properties are anisotropic. While in annealed cubic metals anisotropy appears to be restricted to mechanical properties, hexagonal metals show marked directionality also in electrical, thermal and magnetic properties. There is in these cases rotational symmetry about the hexagonal axis, and the extreme values occur parallel and perpendicular to this axis. If we denote by $p(\alpha)$ the value of the electrical, thermal or magnetic property in a direction making an angle α with the hexagonal axis, and if $p_{\parallel}$ and $p_{\perp}$ are respectively the values of p parallel and perpendicular to the axis, it follows that

$$p(\alpha)=p_{\perp}\sin^2\alpha+p_{\parallel}\cos^2\alpha. \qquad \ldots\ldots(1)$$

Elastic and plastic properties do not conform to this simple relation, nor are the extreme values necessarily in simple crystallographic directions.

For a hexagonal polycrystalline aggregate it is possible in principle to calculate the values to be expected for those properties conforming to equation (1). Crystalline interaction and grain boundary distortion effects render calculations difficult, however, and it is not unusual to ignore such effects. Justification for these omissions rests in the fact that good agreement is found between observed and computed values for electrical and thermal properties of annealed materials.

For a small-grained polycrystalline specimen it is easy to obtain the mean value of any property with rotational symmetry about one direction. Thus in hexagonal metals with a completely random arrangement of crystallite axes, averaging (1) over all orientations leads to

$$\bar{p} = \tfrac{2}{3}p_{\perp} + \tfrac{1}{3}p_{\parallel}. \qquad \ldots\ldots(2)$$

Departure from a random distribution will normally result in directional effects being evident in the observed values of the property p. In such cases preferred orientations in a polycrystalline aggregate can be inferred from measurements of p in three mutually perpendicular directions. Provided the principal values of p are known, the degree of orientation can be estimated in simple cases from equation (1). When complex textures are present the answer is not necessarily unique; however, the information obtained is still a valuable supplement to that deduced from conventional X-ray or optical determinations.

In the work described in this paper magnetic susceptibility determinations have been made on a number of specimens of commercially pure titanium metal, after various mechanical and thermal treatments. Both rolled and compressed specimens have been investigated, and, making use of earlier determinations of texture in heavily rolled titanium (McHargue and Hammond 1953), values have been inferred for the principal susceptibilities of the hexagonal titanium lattice. Using these values estimates have been made of the average degree of preferred orientation at various stages during cold working or annealing of the rolled and compressed metal.

Using titanium compressed to 80% reduction and assuming an approximation to single crystal behaviour for the magnetic properties, the temperature variations of the principal susceptibilities have also been investigated over the range from 25°C to 300°C. The mean value of susceptibility obtained using equation (2) then agrees closely over this temperature interval with that found for samples which are initially magnetically isotropic.

§ 2. Experimental Procedure

2.1. *Balance and Magnet*

Magnetic susceptibilities were measured using the Faraday method and experimental arrangements were sufficiently conventional to require only a minimum description. Specimens of the commercially pure titanium ranged in mass from about 0·08 g to 0·5 g, and forces were determined by an Ainsworth FHM type micro-balance mounted appropriately over the electromagnet. Oil damping was used on the balance, reducing the initial sensitivity of 2 microgrammes to a value of 10 microgrammes per division; this sensitivity was adequate for the accuracy required here.

Specimens were cemented to a fine quartz fibre suspension, and adjusted appropriately in relation to the magnetic field by raising or lowering the balance. No provision was made for evacuating the balance system, but an arrangement of glass tubing around the specimen protected it from external disturbances. The small necessary corrections were applied for magnetic effects due to the quartz suspension and due to the air surrounding the specimen.

The electromagnet was energized from storage batteries, and hand regulation of the magnet current to closely reproducible meter settings was used to control the field. All force determinations and field calibrations were carried out at these same settings. Maximum fields available, at the specimen position, were about 7000 oersteds. To obtain a reasonable volume having a uniform value of $H(dH/dy)$, special pole-pieces were made (Donoghue 1953) and fitted to the magnet. With these pole-pieces values of $H(dH/dy)$ reaching $11{\cdot}5 \times 10^6$ c.g.s. units were attainable.

2.2. *Field Calibration*

Values for the field and field gradient in the present work were obtained in the following manner. A direct calibration was first made using a rotating search coil and fluxmeter to obtain the field H along a vertical line through the centre of the magnet gap. Large scale plots gave provisional values of the field gradients for a number of different magnet currents. Susceptibility determinations were then made using six different specimens of annealed high purity aluminium, each one containing a different amount of ferromagnetic impurity. (Impurity contents varied from 1 to 13 parts per million by weight.) A Honda–Owen plot of apparent susceptibility against $1/H$ was made for each sample, using the H and dH/dy values obtained from the search coil calibration. Consistent deviations from the linear plot of the susceptibility values for *all* the samples at *one* nominal field value can be attributed primarily to an incorrect value of H or of $H(dH/dy)$ at that field. It was first assumed that H was correct and small changes made in the $H(dH/dy)$ value until no consistent deviations from linearity appeared in the $1/H$ plots for the three samples of lowest ferromagnetic content. Any deviations then remaining for the other three samples were reduced by a small change of the H values; this has to be done with discretion to avoid unduly impairing the linearity of the plots for the first three samples. Finally, after these adjustments have been made, it is necessary that the best straight lines through the corrected points should all extrapolate at $1/H=0$ to the same value of susceptibility for each sample.

This 'trial and error' method of reducing systematic deviations arising from inaccuracies in H and dH/dy is simple to carry out and considerably improves the internal consistency of the results. In the measurements on titanium random errors in force determinations appear to account for the major part of the remaining inaccuracy, and it is estimated that the absolute values deduced for susceptibilities should not be in error by more than 1%. These values were in all cases the extrapolated values, obtained from the plot of apparent susceptibility against $1/H$. Field dependence of susceptibility, $d\chi/d(1/H)$, was approximately 10^{-3} c.g.s. units-oersted, corresponding to an impurity content, assumed to be ferromagnetic iron, of about 5 parts per million by weight.

2.3. *Material and Treatment*

A complete analysis of the commercially pure titanium used in this work showed the following weight per cent of elements present: C, 0·074%; N, 0·057%; Fe, 0·15%; Cu, 0·005%; Mg, 0·003%; O, 0·087%; H, 0·007%. Cold working was always carried out at room temperature, either in a compression testing machine or by passing between rollers, and is expressed conventionally as a percentage reduction in thickness. Initial heat treatment or annealing was dependent on the experiment being carried out and is detailed in each case with the relevant susceptibility measurements.

§ 3. Results

3.1. *Susceptibility of Cold-Rolled Titanium*

The starting material for these investigations was a rod of the commercially pure metal, swaged to a diameter of 0·5 inch, then heat treated by annealing in vacuum for four hours at 960°c followed by slow cooling. This particular treatment had been found from preliminary experiments to result in the polycrystalline titanium being magnetically isotropic to within about 0·5%. A number of samples was then cut from the rod after cold-rolling at room temperature to successively increasing reductions. Rolling was always carried out in the same direction and after each reduction the corresponding sample cut from a region near the centre of the cross section. Eight samples were prepared in this way, their masses varying between 0·3 g and 0·15 g and the reduction in thickness extending up to 90%. After etching, each specimen was mounted and its susceptibility measured in three mutually perpendicular directions, perpendicular to the rolling plane and parallel and transverse to the rolling direction. Unless otherwise stated, all susceptibility measurements were made at a temperature of 25°c ± 1°c.

The results of this set of experiments are shown in figure 1. Commencing with the quasi-isotropic material the susceptibility in the rolling plane decreases as reduction proceeds, while perpendicular to the rolling plane there is an increase. This continues up to 60% reduction, beyond which more elaborate changes become evident. The significance of these results in relation to the deformation texture is considered in § 3.3.

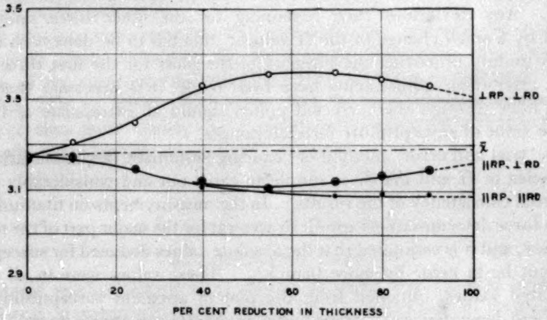


Figure 1. Susceptibility of cold-rolled titanium.

Several x-ray investigations have been made of the texture of heavily rolled titanium, and of the texture when subsequently annealed (see for example Clark 1950, Williams and Eppelsheimer 1953, McHargue and Hammond 1953). The latter authors concluded that the ultimate texture in rolled material was one with the basal plane parallel to the rolling direction but inclined at an angle of 27° to it, and with a [10$\bar{1}$0] direction parallel to the rolling direction. This situation is shown in figure 8 (*a*). McHargue and Hammond further found that annealing at various temperatures up to 820°c resulted in a texture change which could be represented by a rotation of 30° about the normal to the basal plane, the latter still remaining tilted at 27° to the rolling plane; this is shown in figure 8(*b*). This texture should be indistinguishable magnetically from the heavily cold-rolled texture. To investigate this point, susceptibility determinations were made after subjecting the 90% cold-rolled specimen to successive one-hour vacuum anneals at progressively higher temperatures up to 850°c (except at 300°c where a four-hour anneal was given). The results obtained for susceptibilities in the three significant directions are shown in figure 2(*b*). While there is probably a slight progressive fall in two of the susceptibility values as the annealing temperature is raised, this is within the experimental error and the magnetic results conform substantially with the annealing texture found by McHargue and Hammond (1953).

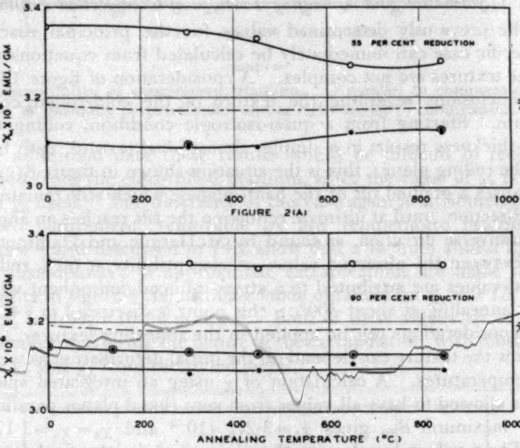


Figure 2. Effect of annealing on susceptibility of cold-rolled titanium. (Significant directions are indicated in the same manner as in figure 1.) (*a*) initial reduction 55%, followed by vacuum anneal. (*b*) initial reduction 90%, followed by vacuum anneal.

It should be noted that McHargue and Hammond's results refer to iodide titanium. From other work (Williams and Eppelsheimer 1953) it appears unlikely that there is any significant difference in textures of iodide and commercially pure titanium.

3.2. *Calculation of Principal Susceptibilities*

Assuming the tilt of the basal plane relative to the rolling plane is known and that the texture is ideal, it is then possible to calculate the principal susceptibilities from the observed values χ_1, χ_2, χ_3. Referring to figure 8(*a*) we have the following relations:

$$\chi_1 = \chi_{\parallel} \cos^2\theta + \chi_{\perp} \sin^2\theta \qquad \ldots\ldots(3)$$

$$\chi_2 = \chi_{\perp} \cos^2\theta + \chi_{\parallel} \sin^2\theta \qquad \ldots\ldots(4)$$

$$\chi_3 = \chi_{\perp}. \qquad \ldots\ldots(5)$$

From a consideration of the results given in figures 1 and 2 (*b*) the most probable values of mass susceptibilities for a completely ideal texture (assumed to be at 100% reduction) are as follows: $\chi_1 = 3{\cdot}29_0 \times 10^{-6}$, $\chi_2 = 3{\cdot}14_2 \times 10^{-6}$, and $\chi_3 = 3{\cdot}08_7 \times 10^{-6}$ e.m.u. g^{-1}. Taking θ to be 27°, and using equations (3) and (5), we obtain $\chi_{\parallel} = 3{\cdot}34_0 \times 10^{-6}$ and $\chi_{\perp} = 3{\cdot}08_7 \times 10^{-6}$ e.m.u. g^{-1}. A check calculation of χ_2 using equation (4) gives a value of $3{\cdot}14_1 \times 10^{-6}$ as compared with the measured value $3{\cdot}14_2 \times 10^{-6}$. Changing θ by 3° would put the calculated value of χ_2 just outside the probable experimental error. From this we conclude that the texture is close to an ideal one and that the results are self-consistent with $\theta = 27°$. The magnetic anisotropy of the commercially pure paramagnetic titanium is thus $\chi_{\parallel} - \chi_{\perp} = 0{\cdot}25_3 \times 10^{-6}$ e.m.u. g^{-1}.

3.3. *Deformation and Annealing Textures of Cold-Rolled Titanium*

Using the previously determined values for the principal susceptibilities, θ in any specific case can immediately be calculated from equations (3) and (4), provided the textures are not complex. A consideration of figure 1 leads to the following conclusions regarding the texture of the cold-rolled commercially pure titanium. Starting from a quasi-isotropic condition, rolling to approximately half-thickness results in a simple, almost ideal texture, with basal planes parallel to the rolling plane; this is the situation shown in figure 8 (*c*). Further reduction causes a gradual tilt of the basal planes, which still remain parallel to the rolling direction, until at ultimate reduction the tilt reaches an angle of about 27° to the tranverse direction, as found by McHargue and Hammond. Small differences between the observed values of susceptibility at 60% reduction and the principal values are attributed to a stress-induced component which can be removed by annealing at about 300°C; this point is discussed in § 4.2.

Similar considerations can be applied to the annealing textures, and figure 2 illustrates how the texture can depend on the initial deformation as well as on the annealing temperatures. A calculation of χ using an integrated space average in which θ is allowed to have all values from zero (basal planes parallel to rolling plane) to a maximum θ_{m}, gives $\chi_1 = 3{\cdot}27_5 \times 10^{-6}$ and $\chi_2 = \chi_3 = 3{\cdot}12_2 \times 10^{-6}$ if $\theta_{\mathrm{m}} = 45°$. These values are close to those observed for the final texture of the material reduced 55% and annealed at 850°C. Thus a spread of tilt of the basal planes of approximately ± 45° from the rolling plane could develop in this case, in contrast to the heavily rolled material where similar annealing produces little change from the initial cold-rolled values of χ.

3.4. *Susceptibility and Deformation Texture of Compressed Titanium*

In the following observations individual samples of material, initially magnetically isotropic with 0·5%, were cold-compressed by successively increasing amounts and the room temperature susceptibilities measured at each stage of compression. A maximum reduction in thickness of 80% was obtained without undue cracking of the specimens. Figure 3 shows the results on one sample, which is typical of the changes observed in the susceptibilities measured parallel and perpendicular to the compression direction.

A progressive increase amounting to nearly 6% is found in the susceptibility measured parallel to the compression axis, and the extrapolated value at ultimate reduction is about $3{\cdot}36_5 \times 10^{-6}$ e.m.u. g^{-1}. In the plane perpendicular to the compression direction the suceptibility is isotropic and, within the limits of error, unchanged up to 80% reduction, the value being $3{\cdot}15_5 \times 10^{-6}$ e.m.u. g^{-1}.

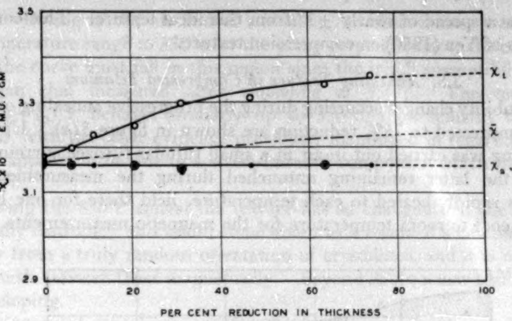


FIGURE - 3

Figure 3. Susceptibility of compressed titanium. ○ parallel to compression direction. ⊙, ● mutually perpendicular, and at right angles to compression.

Without additional data these results would be difficult to reconcile with those obtained from the cold-rolled titanium, since the value of χ_1 is apparently greater than $\chi_{\parallel}$ itself. The difference in this case again is attributed to a positive stress-induced component, removable by low temperature heating. Results given in figure 4 and discussed in § 4.2 show that a vacuum anneal at 300°c will remove the discrepancy; if approximate extrapolations are made through the annealed points in figure 4 the ultimate value obtained for χ_1 is $3{\cdot}32_5 \times 10^{-6}$, and for χ_2 and χ_3 is $3{\cdot}09_0 \times 10^{-6}$ e.m.u. g^{-1}. Both of these values are in sufficiently good agreement, considering the nature of the extrapolation, with those previously deduced for $\chi_{\parallel}$ and $\chi_{\perp}$ from the rolled metal.

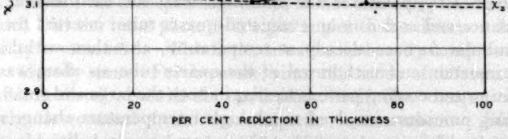


Figure 4. Effect of annealing at intermediate stages of compression. At A annealed four hours at 300°c, followed by compression to B, then further annealed two hours at 300°c

It is concluded that the deformation texture of highly compressed titanium is a simple one with the basal plane of the lattice normal to the axis of compression, approximately as shown in figure 8 (*c*). The results of the present experiments would allow a spread of nearly ± 3° from this ideal texture. This conforms to the findings of Yen (1950) on compression texture.

3.5. *Annealing Texture in Compressed Titanium*

Susceptibility changes occurring during the progressive annealing of a sample initially compressed to 75% reduction are shown in figure 5 (*a*). Up to 350°c the annealing was carried out in air in a small tubular furnace surrounding the specimen, the latter remaining untouched during the measurements. The sample was rapidly heated to each temperature, held there for one hour, and allowed to cool to room temperature for the magnetic measurements.

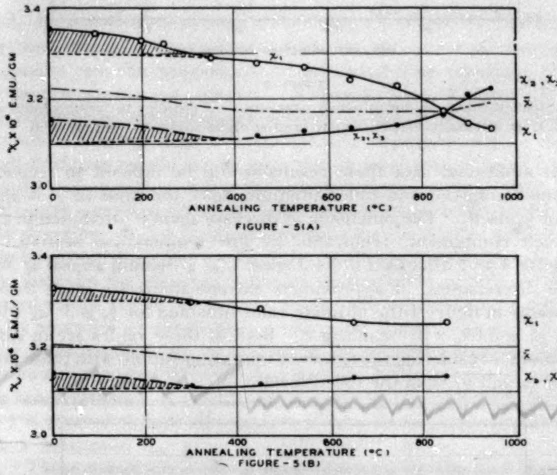


Figure 5. Effect of annealing : ○ measured parallel to compression direction, ● measured in plane perpendicular to compression direction. (*a*) initial compression 75%. (*b*) effective initial compression 25%. Shaded areas represent regions where stress-induced effects are probably operative.

At the higher temperatures the specimen was, for each anneal, removed from the balance and sealed in an evacuated quartz tube, inserted for one hour in a large tubular furnace already at temperature, and then withdrawn. By appropriate insertion and withdrawal of the quartz tube an effort was made to keep the heating and cooling periods similar in both the large and small furnaces. This annealing procedure involved much higher temperature change rates than those used previously.

The susceptibility χ_1 measured parallel to the original compression axis falls as the sample is successively annealed, from a value of $3{\cdot}35_5 \times 10^{-6}$ to about $3{\cdot}12_5 \times 10^{-6}$ after annealing at 950°c. Although the experimental error is such that the decrease could be regarded as a smoothly progressive one, we have preferred to give greater significance to the trend of individual points so that the curve appears with an enhanced drop down to 350°c and thereafter a slight flattening. Owing to the method of annealing, no observations were available in the temperature range to 350°c for the plane perpendicular to the compression axis, but the curve must fall in this region since the initial susceptibility value is higher than that measured after annealing at 450°c. The mean value, $\bar{\chi}=(\chi_1+\chi_2+\chi_3)/3$, behaves similarly up to 350°c, then remains constant to 850°c, and, with the other evidence previously mentioned in regard to figure 4, this again appears to be due to a stress-induced component which has annealed out at 350°c.

Following the 850°c anneal the texture has so changed that the material is magnetically isotropic; this could come about as a result of a complex multiple texture or from a truly random orientation of crystallites, and it is not possible to distinguish between these magnetically. Beyond 950°c a new texture appears to be developing.

Another type of annealing curve for an initially compressed material is shown in figure 5 (*b*). Here the specimen had an effective initial reduction of 25%. In this case there is no suggestion of the material becoming completely isotropic, although the plane normal to the compression axis again remains so. There is also an initial decrease in the mean susceptibility, as previously observed. Assuming that a single texture is present in this case after the first anneal at 850°c, and taking $\chi_1=3{\cdot}24_5\times10^{-6}$, we obtain from equation (3) a value of 38° for the angle θ. Using equation (4) with this value of θ, and allowing for the random arrangement about the compression axis, the calculated susceptibility in the isotropic plane is then $3{\cdot}13_8\times10^{-6}$. This is within experimental accuracy of the observed value, $3{\cdot}12_7\times10^{-6}$. Thus the texture in this case could be one where the basal planes make an angle of about 38° with the plane normal to the compression axis and are randomly arranged about the axis. Alternatively, the texture may be such that θ is uniformly spread from zero (basal plane perpendicular to compression axis) up to a maximum $\theta_{\rm m}$. In this case an integrated space average, using $\theta_{\rm m}=55°$, gives $3{\cdot}25_0\times10^{-6}$ for χ_1, and $3{\cdot}13_2\times10^{-6}$ for χ_2 and χ_3. This type of texture would therefore equally well satisfy the experimental results.

3.6. *Ambiguity in Texture Determinations*

As the above calculations illustrate, a measure of ambiguity must usually exist in the texture determination; it is not in general possible to calculate a unique texture which will conform to the magnetic observations for any particular case. However, if one of the susceptibility values is the same as $\chi_{\parallel}$ the texture can then only be such that the basal plane is at right angles to the direction in which this susceptibility is measured. This is the situation in heavily compressed material, as we have already seen. Furthermore, since this mode of deformation necessitates a random distribution in the basal plane itself, the 'compression texture' is uniquely determined from the magnetic measurements. Similarly, the texture in cold-rolled material at about 50 to 60% reduction is like the compression texture, though here the basal plane may have a directionality which could not be detected magnetically.

Annealing textures are subject to the same indeterminacies, and again one can either postulate double textures or make calculations on the assumption of smoothly progressive texture changes. The latter was done in those cases already discussed.

Even though the complete specifications of the degree of preferred orientation may be uncertain, it is still possible to make some qualitative observations of value in regard to the actual working of the metal, using as a basis the observed annealing changes in susceptibility. For many purposes an isotropic texture is mechanically desirable in a metal which is being formed, and from the present results it is clear that, after an *intermediate* degree of rolling or compression, an anneal in the 600°c range will tend to return the material to a more isotropic condition. After *heavy* rolling, on the other hand, the texture still maintains its high degree of orientation irrespective of the annealing, up to 850°c at any rate. Heavily compressed material, after a quenching type of anneal, also tends to return to a magnetically isotropic state, as shown in figure 5 (*a*). The original swaged material becomes isotropic after annealing for several hours above the transformation temperature. Thus in these cases, even if the texture details are not certain, the magnetic observations indicate whether one particular type of cold working or heat treatment will give a texture which is mechanically preferable to that produced by another treatment.

3.7. *Susceptibility Variation in Material of initially unknown Texture*

As a final example of more elaborate magnetic changes brought about by texture variations resulting from cold work, figure 6 gives susceptibilities in three significant directions for samples cut from a swaged rod and then subjected to compression. The direction of compression relative to the axis of the rod (and hence the swaging direction) is shown in the diagrams. Two features of these curves are evident. First, at relatively large reductions the susceptibility χ_1, parallel to the compression axis, increases and tends towards a maximum

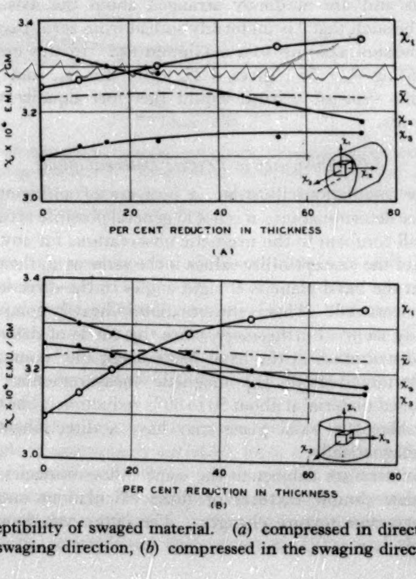


Figure 6. Susceptibility of swaged material. (*a*) compressed in direction at right angles to swaging direction, (*b*) compressed in the swaging direction.

value in a manner similar to that seen in figure 3. Secondly, the susceptibilities χ_2 and χ_3 measured in the plane perpendicular to the compression axis both appear to approach the same ultimate value, again similar to figure 3. The material is in fact acquiring the 'compression texture' previously discussed.

Considering the starting values of the susceptibilities, it is clear that the swaged rod initially has a texture with all the basal planes approximately parallel to the rod axis. The susceptibility measured in this direction is $3{\cdot}09_0 \times 10^{-6}$ e.m.u. g^{-1}, indicating an absence of any stress-induced effect; in this respect the situation is similar to that in the case of heavily rolled material. Unequal susceptibilities in the plane normal to the axis are to be expected since both specimens were cut from a position near to the surface of the rod. This makes an exact calculation of crystal axes in the plane impossible; however, a spread of approximately ±40° about the direction of χ_2 in each case would be in fair agreement with the observed susceptibilities in the plane normal to the rod axis. These observations suggest that, if any stress-induced effects are first removed by a low temperature anneal, a reasonable estimate of texture should be possible in all cases where three independent susceptibility values can be measured.

3.8. *Temperature Variation of Mean Susceptibility*

The temperature variation of the average susceptibility has been measured by several workers, both for commercially pure and for iodide titanium (Klemm 1939, Squire and Kaufmann 1941, Denney 1955). Room temperature values lie between $3{\cdot}18 \times 10^{-6}$ and $3{\cdot}31 \times 10^{-6}$ e.m.u. g^{-1}, with a temperature coefficient of $0{\cdot}0012 \times 10^{-6}$ e.m.u. g^{-1} deg^{-1}. There appear to be no recorded measurements on single crystals of titanium.

The present results suggest that close approximation to single crystal values should follow from measurements on heavily compressed material, and 'principal susceptibilities' for commercially pure titanium over the temperature range from 25°c to 300°c have been obtained in this way. Starting with quasi-isotropic material a reduction in thickness of 80% was made by straight compression. Previous measurements show that this will result in a texture with the hexagonal axes of the crystallites all very nearly parallel to the compression direction. The susceptibility in this direction should therefore be close to the value of $\chi_{\parallel}$, although probably modified by the stress-induced increase already noted after compression. Similarly, measurement in the plane normal to the compression direction will give approximately $\chi_{\perp}$. We shall call the measured values of these susceptibilities χ_{max} and χ_{min} respectively.

Following compression the sample was attached to the balance suspension, oriented to measure χ_{max}, and surrounded by a small tubular furnace mounted between the magnet pole-pieces. A direct reading potentiometer was used to measure temperatures, connected to a copper–constantan thermocouple held centrally in the furnace and about 5 mm below the specimen. It is estimated that the recorded temperatures are within ±3°c of the specimen temperature over the range to 500°c. The initial value measured for χ_{max} at 25°c was $3{\cdot}37_0 \times 10^{-6}$ e.m.u. g^{-1}, in close agreement with previously observed values for heavily compressed material. After heating for three hours at 300°c this had fallen to $3{\cdot}32_7 \times 10^{-6}$ e.m.u. g^{-1} and was thereafter constant. Susceptibilities were measured in this orientation for a number of temperatures up to 300°c, with the results shown in figure 7. The specimen was not touched during these observations, and the small corrections for the magnetic effect of the suspension and the surrounding air were applied as before; in the present instance the mass of the sample was approximately 0·1 g and the total corrections amounted to nearly 2% of the measured force.

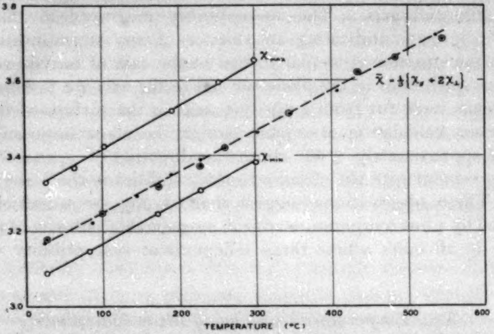


Figure 7. Temperature variation of ' principal susceptibilities ' of titanium. — — — — average value $\bar{\chi}$, deduced from principal susceptibilities. ⊙, ● measured susceptibilities of isotropic specimens.

Similar observations were made with the specimen oriented to measure χ_{min}. The value at 25°c, after the 300°c anneal, was $3{\cdot}08_7 \times 10^{-6}$ e.m.u. g^{-1}, again about identical with that computed for $\chi_{\perp}$ from the cold-rolled material. The temperature coefficients for χ_{max} and χ_{min} are very nearly equal over the measured temperature interval, the average being $0{\cdot}0011 \times 10^{-6}$ e.m.u. g^{-1} deg^{-1}. This value agrees well with that found by Denney (1955).

After the work described here was completed a piece of large-grained iodide titanium was very kindly made available by Watertown Arsenal, Massachusetts, U.S.A. From this a single crystal specimen of mass about 0·3 g was cut and the direction of the hexagonal axis determined by x-ray diffraction. Magnetic measurements on this crystal provided a direct check on the deductions already made from the results on the cold-worked metal. The principal susceptibilities so determined are in good agreement with the values previously deduced. It is also apparent that the magnetic susceptibilities, at room temperature, of iodide and of commercially pure titanium are closely similar. Using the results of figure 4 to allow for slight changes in susceptibility which might be expected if the texture were completely ideal at the ultimate reduction, the table gives computed values of principal susceptibilities obtained from the various observations.

	(1)	(2)	(3)
$\chi_{\parallel}$	$3{\cdot}34_0 \times 10^{-6}$	$3{\cdot}33_7 \times 10^{-6}$	$3{\cdot}35_0 \times 10^{-6}$
$\chi_{\perp}$	$3{\cdot}08_7 \times 10^{-6}$	$3{\cdot}08_2 \times 10^{-6}$	$3{\cdot}07_0 \times 10^{-6}$
$d\chi/dT$	—	$0{\cdot}0011_3 \times 10^{-6}$	—
Anisotropy, $(\chi_{\parallel} - \chi_{\perp})$	$0{\cdot}25_3 \times 10^{-6}$	$0{\cdot}25_5 \times 10^{-6}$	$0{\cdot}28_0 \times 10^{-6}$
$\bar{\chi}$	$3{\cdot}17_3 \times 10^{-6}$	$3{\cdot}16_7 \times 10^{-6}$	$3{\cdot}16_3 \times 10^{-6}$

Temperature 25°c ; (1) calculated from known texture of cold-rolled metal ; (2) calculated from results for 80% compressed metal ; (3) observed values from single crystal.

With a completely random arrangement of the crystallites in a sample of a hexagonal metal it follows from equation (2) that the average susceptibility would be given by $\bar{\chi}=(\chi_{\parallel}+2\chi_{\perp})/3$. In figure 7 are also shown the measured susceptibilities for two other samples, over the temperature range up to 540°c, after they had been initially made magnetically isotropic. The broken line gives $\bar{\chi}$ calculated from the deduced values of $\chi_{\parallel}$ and $\chi_{\perp}$. It will be seen that the observed susceptibilities closely follow the calculated average value.

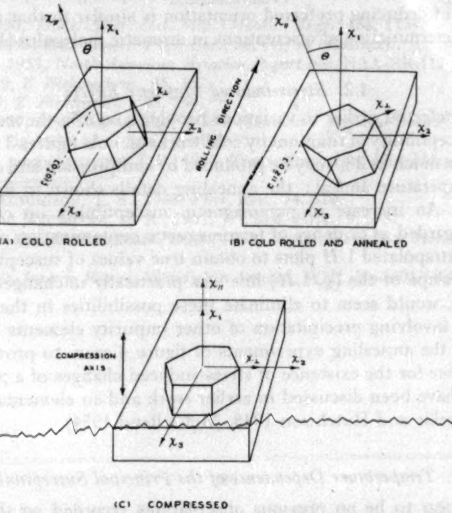


Figure 8. Probable textures in heavily cold worked titanium.

The best value of average susceptibility obtained from the present work with commercially pure titanium is $3{\cdot}17\times10^{-6}(\pm0{\cdot}03\times10^{-6})$ e.m.u. g^{-1} at 25°c. This differs appreciably from the figure of $3{\cdot}31\times10^{-6}(\pm0{\cdot}02\times10^{-6})$ given by Denney (1955) for a temperature of 23°c, but is in close agreement with the earlier results of Klemm (1939) and of Squire and Kaufmann (1941). The presence of oxygen, which is the most easily acquired impurity, should have negligible effect on the susceptibility for the amounts existent in the material used here (see Ehrlich 1939). Oxygen contents are not quoted in the other susceptibility determinations.

§ 4. Discussion

4.1. *Preferred Orientations*

Potentially the most useful feature in the present results is probably their application to texture determinations in the cold worked or annealed metal during fabrication processes. To obtain a complete estimate of such texture it is normally necessary to carry out a detailed analysis of the x-ray or neutron diffraction from a sample of the metal; these analyses are often lengthy. Within certain limits, however, rapid estimates can be obtained from magnetic measurements alone.

Texture determinations are only possible, of course, when an appreciable magnetic anisotropy exists in the metal. All cubic metals in their normal condition are thus immediately eliminated in this connection. Although titanium is paramagnetic, any non-cubic diamagnetic metal could be equally well investigated, provided only that the anisotropy were sufficiently large. It will be clear that this method of deducing preferred orientation is similar to that which has been used in the determination of orientations in aromatic molecules (Lonsdale 1937).

4.2. *Stress-induced Magnetic Effects*

We have referred earlier to variations brought about in the *mean* value of the magnetic susceptibility of titanium by cold working. As figures 3 and 4 illustrate, increases of as much as 2% may be produced by compression and can be removed by a low temperature anneal; the annealing details shown in figure 4 are self-explanatory. An increase in paramagnetic susceptibility on cold working is frequently regarded as evidence of ferromagnetic contamination or precipitation. The use of extrapolated $1/H$ plots to obtain true values of susceptibility, and the fact that the slope of the $(\chi, 1/H)$ line was practically unchanged whatever the cold working, would seem to eliminate these possibilities in the present work. Explanations involving precipitation of other impurity elements appear equally unlikely, and the annealing experiments of figure 4 seem to provide fairly conclusive evidence for the existence of stress-induced changes of a primary nature. Such effects have been discussed in earlier work and an elementary explanation suggested (Reekie and Hutchison 1948, McClelland 1954).

4.3. *Temperature Dependence of the Principal Susceptibilities*

There appear to be no previous observations recorded on the variation of principal susceptibilities with temperature, although Squire and Kaufmann (1941) investigated the *average* susceptibility as a function of temperature. Their results covered the range from about -200°c up into the β region at 1200°c and showed a positive, nearly linear, temperature coefficient for the measured susceptibility over a large part of the interval lying in the α region. Using initially isotropic samples the present results for $\bar{\chi}$ agree well, both as regards susceptibility values and temperature coefficient, with Squire and Kaufmann's observations over the temperature range in common.

A point of significance emerging from the present results is the fact that the principal susceptibilities themselves appear to have identical temperature coefficients over the range from 25°c to 300°c. The validity of this deduction is dependent, of course, upon identifying $\chi_{\max}$ and $\chi_{\min}$ of figure 7 with $\chi_{\parallel}$ and $\chi_{\perp}$ respectively. From our limited single crystal measurements this appears justified. It is unfortunate that the temperature range over which the principal susceptibility values could be measured was restricted to a rather narrow interval in the present case, and it would be very desirable to extend these measurements. However, over this range, the magnetic anisotropy $\chi_{\parallel} - \chi_{\perp}$ is independent of temperature.

ACKNOWLEDGMENTS

This work is published with the approval of the Deputy Minister of Mines and Technical Surveys, Ottawa. The authors are indebted to several members of the Physical Metallurgy Division for helpful discussions on metallurgical points relating to titanium. The assistance of Mr. J. R. Rowland in determining axial directions in the single crystal used in the latter part of the work, and of Mr. K. S. Milliken for help in deducing space average values, is particularly appreciated.

REFERENCES

CLARK, H. T., 1950, *Trans. Amer. Inst. Min.* (*Metall.*) *Engrs*, **188,** 1154.

DENNEY, J. M., 1955, *Bull. Amer. Phys. Soc.*, **30,** (4), 14 (abstract).

DONOGHUE, J. J., 1953, *North American Aviation Report* No. NAA-SR-117.

EHRLICH, P., 1939, *Z. Elektrochem.*, **45,** 362.

KLEMM, L., 1939, *Z. Elektrochem.*, **45,** 354.

LONSDALE, K., 1937, *Rep. Progr. Phys.*, **4,** 368. (London : Physical Society).

MCCLELLAND, J. D., 1954, *Acta Met.*, **2,** 406.

MCHARGUE, C. J., and HAMMOND, J. P., 1953, *Trans. Amer. Inst. Min.* (*Metall.*) *Engrs*, **197,** 57.

REEKIE, J., and HUTCHISON, T. S., 1948, *Phys. Rev.*, **74,** 610.

SQUIRE, C. F., and KAUFMANN, A. R., 1941, *J. Chem. Phys.*, **9,** 673.

WILLIAMS, D. N., and EPPELSHEIMER, D. S., 1953, *Trans. Amer. Inst. Min.* (*Metall.*) *Engrs*, **197,** 1378.

YEN, M. K., 1950, *Interim Report*, Watertown Arsenal, WHL No. 401/14-4.

The Lattice Spacings of Thin Electrodeposits of Cobalt and Nickel on a Copper Single Crystal

By R. C. NEWMAN †
Department of Physics, Imperial College, London

Communicated by M. Blackman ; *MS. received* 14*th October* 1955, *and in final form* 12*th December* 1955

Abstract. An electron diffraction investigation has been made of the lattice spacings of cobalt and nickel electrodeposited on to a copper single crystal. By the use of radioactive cobalt to measure the thickness of electrodeposits, it is shown that no continuous metal deposition occurs at cathode current densities less than $90\,\mu\text{A}\,\text{cm}^{-2}$. At current densities greater than this value, oriented cobalt is deposited. The results for nickel are similar and the critical current density is $100\,\mu\text{A}\,\text{cm}^{-2}$. The results show that where cobalt and nickel are deposited they have their normal lattice spacings. Previous claims that these metals form pseudomorphic layers have been re-interpreted without this assumption.

§ 1. Introduction

The electron diffraction technique is particularly suitable for the investigation of the structure of thin layers of electrodeposited metals, and many workers have studied how this structure is influenced by the substrate metal, and by the rate of deposition. When a single crystal substrate is used, the deposit often forms an oriented layer. Thus, Thomson (1931) found that silver, electrodeposited on to a copper single crystal, grew with its lattice parallel to the copper lattice, and Finch *et al.* (1947) found that thin electrodeposits of several metals on single crystals of copper and iron formed oriented layers. Cochrane (1936) investigated electrodeposits, made at various current densities, of several metals on copper single crystals. The surprising results were that thin layers (up to about 100 Å), of both cobalt and nickel, deposited at current densities of $4\,\mu\text{A}\,\text{cm}^{-2}$ and $25\,\mu\text{A}\,\text{cm}^{-2}$ respectively, formed oriented layers which possessed the lattice spacing of copper instead of their normal values, the accuracy being quoted to better than 1%. At a current density of $50\,\text{mA}\,\text{cm}^{-2}$ a mixture of randomly oriented cubic and hexagonal cobalt was deposited. Nickel deposits at current densities greater than $300\,\mu\text{A}\,\text{cm}^{-2}$ were randomly oriented and thin layers were again stated to have the lattice spacing of copper. The basis for this claim was that the diffraction rings due to the nickel passed directly through the spots in the pattern, which Cochrane attributed to the copper substrate.

This effect of the lattice spacing of a thin deposit changing to fit exactly with that of the substrate has been called 'basal-plane pseudomorphism' by Finch and Quarrell (1933). There is, however, a fundamental objection to the occurrence of this effect in the case of a randomly oriented deposit on a single crystal substrate. If the interaction between the substrate and the deposit is insufficient to produce an oriented overgrowth, then there is no apparent reason why the lattice spacing of the deposit should be in any way modified by the substrate lattice. In the case of oriented deposits, it is difficult to see how the lattice spacing of the deposit can be modified up to thicknesses of 100 Å, as was claimed, since the range of the forces from the substrate atoms extend only to the first few atomic layers of the deposit.

---

† Now at A.E.I. Research Laboratory, Aldermaston.

Two other cases of pseudomorphism have been reported by Finch and Quarrell (1933). These pertained to the growth of zinc oxide on zinc and aluminium evaporated on to platinum. The interpretation of these experiments is now open to considerable doubt, however, as a result of further experiments performed by Raether (1950), Lucas (1951) and Shishakov (1952). Thus the results concerning the formation of pseudomorphic deposits reported by Cochrane are the only ones for which an alternative interpretation has not so far been given. Since the concept of pseudomorphism has gained wide currency, and has been used for example by Frank and Van der Merwe (1949) in their theory of orientation, it seems most desirable that Cochrane's work be repeated. Electrodeposits of cobalt and nickel on a copper single crystal have therefore been re-examined in detail, particular attention being paid to the lattice spacings of the deposits.

§ 2. EXPERIMENTAL PROCEDURE

In all the experiments the etched (110) faces of copper single crystals were used. One crystal was cut from the rod originally used by Cochrane and the others were cut from rods of Johnson, Matthey 'spec-pure' copper single crystals. The crystals were polished on (0000) emery and then etched for ten minutes in a 10% ammonium persulphate solution, washed well in distilled water, alcohol and finally benzene before they were examined in the electron diffraction camera. The diffraction pattern obtained from a typical specimen is shown in figure 1 (Plate) and is seen to consist of the usual cross-grating of spots due to the copper. together with some faint rings due to randomly oriented cupric oxide. This randomly oriented oxide was presumably formed during the etching process or in the washing of the copper specimen. Since oriented deposits of nickel and cobalt have been found (see § 3) it must be concluded that the oxide was dissolved by the electrolyte used in the plating cells.

The plating cell consisted of a small dish with a flat disc of the metal to be deposited as anode. The metals used for the anodes were of high purity and 'Analar' salts were used for the electrolytes. The composition of the electrolytes was (*a*) crystalline cobalt sulphate 300 g l.$^{-1}$, sodium chloride 3 g l.$^{-1}$ and boric acid 6 g l.$^{-1}$ and (*b*) crystalline nickel sulphate 300 g l.$^{-1}$, sodium chloride 3·1 g l.$^{-1}$ and boric acid 6·2 g l.$^{-1}$ for the cobalt and nickel baths respectively; these were the same composition as in Cochrane's experiments. The etched face of the copper crystal was immersed just below the surface of the electrolyte during deposition so that only the freshly etched copper came into contact with the solution. All deposition was carried out at room temperature without agitation of the solution, after which the specimen was again washed well in distilled water, alcohol and finally benzene.

Cochrane found that the structure of his deposited metals depended upon the cathode current density used for the deposition. To test whether this effect was due to the characteristics of the plating cells, their properties have been examined over the range of current densities used by Cochrane.

The cathode current density was measured as a function of the potential across the cell, the curve obtained for the cobalt bath being shown in figure 2. A similar curve was obtained for the nickel cell, the values of V_0 and V_1 being $-0{\cdot}1$ v and $+0{\cdot}55$ v respectively. These curves were only obtained when special care was taken to clean the anode of the plating cell. Repeatable results were obtained if the anode metal was first polished and then lightly etched in nitric acid, followed by a thorough wash in distilled water before it was transferred to the plating cell. Without this procedure the normal e.m.f. of the cell varied with time and the polarity was even reversed in some cases.

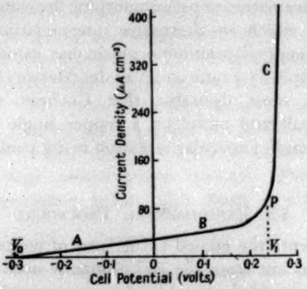


Figure 2. The relationship between current density and the applied cell potential for the cobalt bath.

According to electrochemical data (see Glasstone 1945) metal deposition would only be expected to occur in region C. In regions A and B any cathode current is carried by hydrogen ions. The potential V_1 corresponding to the point P is known as the overpotential and is the sum of two overpotentials V_2 and V_3 set up at the anode and cathode respectively. V_2 and V_3 depend on the surface state of the metal electrodes and hence if the anode is not thoroughly clean, repeatable cell conditions are not expected. Thus, no metal deposition would be expected below the critical current densities of $90\,\mu\text{A cm}^{-2}$ and $100\,\mu\text{A cm}^{-2}$ which correspond to the values of V_1 for cobalt and nickel respectively when a cleaned anode is used in the plating cell. Hence, Faraday's law cannot be used to estimate the thickness of deposits except in region C of the cell characteristics. Since it was vital to know the thickness of any deposit, an independent method of measurement had to be employed.

A large number of radioactive isotopes is now available and it is possible to use tracer methods for thickness measurements in thin films. ^{60}Co is a suitable isotope for this purpose and has therefore been used in the present work. ^{60}Co disintegrates into ^{60}Ni with the emission of 1·2 MeV γ-ray and an electron of 0·308 MeV maximum energy. The active cobalt sulphate solution which contained the correct proportions of non-active sodium chloride and boric acid, had a specific activity of $1{\cdot}2\,\text{mc}\,\text{g}^{-1}$. The β-rays were counted with a G.E.C. type GM4 counter which was enclosed in a lead chamber to reduce the background counting rate; this was about 15 counts per minute. The specimen to be counted was mounted close up to a brass aperture $\frac{1}{8}$ in. diameter immediately above the counter window. In this way, the effective area of all specimens was the same and the thickness of any deposit was proportional to its counting rate, for the thin layers investigated.

To obtain an absolute value of the thickness from the counting rate, it was necessary to carry out some form of calibration experiment. This was effected by the deposition of 1000 Å of active cobalt according to Faraday's law at a current density of 1·6 mA cm^{-2} which is in region C of the deposition curve and where Faraday's law is expected to hold reasonably well. The sensitivity was thus found to be 2 counts min^{-1} Å^{-1} which agreed well with further calibrations at other current densities and with different thicknesses of cobalt deposited in region C of the cell characteristic. It also agreed fairly well with the value calculated from the geometry of the arrangement, the specific activity of the cobalt and the efficiency of the counter. If gas evolution did occur during deposition in the calibration experiments, then the sensitivity would be greater than that quoted above and all thicknesses given below would be smaller than stated. No corrections were made to the counting rates due to the decay of the cobalt since the half life of the isotope is 5·3 years, and the experiments in which the active cobalt solution was used only extended over a period of about two months. The counting was continued for each specimen until about 1000 counts had been registered. When the effect of the background counting rate is considered, this means that thick deposits of 100 Å were measured with accuracy to about ± 3%, while very thin deposits of 1 Å in thickness were only measured with accuracy to ± 30%. The accuracy of all measurements is also subject to any systematic errors in the calibration experiments which may have been as large as 20%.

Unfortunately no corresponding radioactive nickel isotope was available to measure the thickness of the nickel deposits. However, since the characteristics of the nickel plating cell are similar to those of the cobalt, it is inferred that similar processes occur in both cells in corresponding regions of the characteristics.

The lattice spacings of specimens were determined by comparison with a transmission specimen of thallium chloride on Formvar used as a standard (see Boswell 1950). The maximum possible error of these measurements was about 0·3%, and hence the difference between the lattice spacings of a deposit of either cobalt or nickel and the copper substrate was readily distinguishable.

§ 3. RESULTS

3.1. *Cobalt Deposits*

Current was passed for various periods at current densities of 4 μA cm^{-2}, 20 μA cm^{-2}, 40 μA cm^{-2} and 80 μA cm^{-2}, which would have produced cobalt layers of up to several hundred ångströms if Faraday's law held under these conditions. In addition, the copper crystal was also immersed in the plating solution for periods of up to 17 hours without any current being passed. The thickness of the cobalt layer on the copper surface after these treatments, was always about 1–2 Å, as determined by the radioactive tracer technique, and was independent of the period for which the current was passed.

The diffraction patterns obtained from these specimens, however, showed new features compared with the original pattern from the freshly etched copper. At the copper [$\bar{1}$10] azimuth there were strong streaks joining spots along ⟨111⟩

directions, extra spots due to twinning on octahedral planes and two sets of rings as shown in figure 3 (Plate). The first set of rings was due to cupric oxide while the second set which was due to a face-centred cubic substance passed directly through the spots in the pattern. At the [001] azimuth some of the main spots were split into four subsidiary spots and the two sets of rings were again visible. These patterns are similar to those obtained by Cochrane who explained them as due to repeated twinning. (Similar patterns have also been obtained by Kirchner and Lassen (1935) for silver deposits on rock salt.)

At current densities of 40 μA cm^{-2} and 80 μA cm^{-2} the intensity of these new features increased with increasing periods of immersion of the copper crystal in the cobalt plating solution, although the actual thickness of cobalt present remained constant at about 1–2 Å. The lattice spacing was measured for some of these specimens and all gave the normal value for copper.

At a current density of 160 μA cm^{-2} cobalt metal appeared to deposit approximately according to Faraday's law and a layer of 150 Å in thickness could be detected by visual inspection of the specimen. Deposits of 100 Å, 200 Å and 400 Å thick all gave similar diffraction patterns except that the spots due to the copper became weaker as the thickness of cobalt was increased. Besides these spots, there were also spots due to both the face-centred cubic and the hexagonal modifications of cobalt. The cubic cobalt was oriented parallel to the copper and there was some twinning on octahedral planes giving rise to extra spots at the copper $[\bar{1}10]$ azimuth. The orientation of the hexagonal cobalt was such that the (001) planes of the cobalt were parallel to $\{111\}$ planes of copper, with $\langle 100 \rangle$ cobalt directions parallel to $\langle 110 \rangle$ copper directions. The growth which occurred on copper $(\bar{1}11)$ planes perpendicular to the surface gave rise to extra spots at the copper $[\bar{1}11]$ azimuth since this was also the cobalt [001] azimuth (see figure 4 (Plate)). The hexagonal cobalt which grew on the copper (111) and $(11\bar{1})$ planes gave rise to extra spots at the copper $[\bar{1}10]$ azimuth, as this was also the cobalt [100] azimuth. In all cases, spots due to cobalt were quite distinct from those due to copper and measurements showed that the cobalt had its normal lattice spacing.

At a current density of 320 μA cm^{-2}, the patterns obtained were similar to those described for a current density of 160 μA cm^{-2} except that there were some weak rings present due to cubic and hexagonal cobalt. There were also some extra spots from another orientation of hexagonal cobalt with the (001) cobalt plane parallel to the (001) copper plane and the [210] cobalt direction parallel to the [110] copper direction. Some spots due to cobalt at the copper $[\bar{1}12]$ azimuth could not be explained by the above orientations. They were, however, Bragg spots since the rings due to the randomly oriented cobalt passed directly through them.

At higher current densities of 800 μA cm^{-2} and 4 mA cm^{-2} the patterns obtained were similar to those found at a current density of 320 μA cm^{-2}, except that the proportion of randomly oriented cobalt increased somewhat.

All of these results were obtained with a cleaned anode in the plating cell. In addition, further experiments were carried out at current densities lower than the critical value of 90 μA cm^{-2} with an uncleaned anode. Cochrane (private communication) has stated that he found deposition under these conditions as deduced from an observed blackening of the copper surface. No such deposition was found in the present experiments, however, despite several attempts to find the effect. As stated in § 2, the electrolytic cell is not running under reproducible conditions in this case and the present conditions may not therefore have been identical with those used by Cochrane (1936).

3.2. *Nickel Deposits*

Current was passed for various periods at current densities of 25 μA cm^{-2}, 40 μA cm^{-2} and 80 μA cm^{-2} and the copper crystal was also immersed in the plating solution for periods of up to 16 hours without current being passed. None of these specimens showed any surface colour change although the thickness of the deposits should have been several hundred ångströms according to Faraday's law. It seems likely that only 1–2 Å of nickel was deposited in view of the cobalt results. The diffraction patterns from these specimens were indistinguishable from those obtained from specimens treated in the cobalt plating cell at the very low current densities (less than 90 μA cm^{-2}). The lattice spacing of the specimens was found to be definitely that of copper in all cases.

At current densities of 300 μA cm^{-2} and 600 μA cm^{-2}, which are in region C of the cell characteristic, nickel was deposited. A layer of 150 Å in thickness was observable by visual inspection of the copper surface. Deposits of 50 Å, 100 Å, 150 Å, 200 Å, 250 Å and 350 Å were made and the diffraction patterns all showed similar effects. Figure 5 (Plate) shows the pattern obtained at the copper [$\bar{1}$10] azimuth with a nickel deposit of 100 Å. It is seen that there are two sets of spots due to copper and nickel with its normal lattice spacing. There is also some twinning in the nickel deposit and a certain amount of randomly oriented nickel. As the nickel thickness was increased, the pattern from the copper gradually became weaker and the intensity of the rings due to randomly oriented nickel increased.

Additional experiments were carried out in which an uncleaned nickel anode was used in the plating cell. In some cases the surface of the copper became darkened when current was passed at the rate of 40 μA cm^{-2} for long periods. It was not possible to measure the thickness of these deposits but they were very much thinner than the values calculated according to Faraday's law. When the specimens were examined with an optical microscope the surface was seen to be covered with black specks while the rest of the surface appeared to be clean copper. The diffraction patterns obtained were identical in character to those described from specimens treated at low current densities in the plating cells under repeatable conditions and where only 1–2 Å of cobalt were deposited. The lattice spacing of some of these specimens was measured very carefully and was found to be that of copper in all cases. These results are presumably similar to those found by Cochrane for nickel deposition carried out at very low current densities.

§ 4. SUMMARY AND DISCUSSION

The first part of this section is devoted to a summary and discussion of the results obtained in the present work when a cleaned anode was used in the plating cell. In the second part, the results obtained by Cochrane (1936) are discussed in relation to these experiments and to the experiments in which an uncleaned anode was used.

By the use of radioactive cobalt to measure the thickness of deposits, it has been shown that cobalt does not deposit on to a copper single crystal according to Faraday's law when current densities less than 90 μA cm^{-2} are used. The actual thickness of cobalt present after current had been passed at these very low current densities was about 1–2 Å in all cases and was independent of the 'deposition' period. When the copper crystal was examined by electron diffraction, the patterns showed new features (as described in § 3) compared with the original pattern from the freshly etched copper; the intensity of these new features increased with increasing periods of immersion of the specimen in the plating cell at current densities of 40 μA cm^{-2} and 80 μA cm^{-2}. The measured lattice spacing of all specimens was found to be that of copper.

The possible effect on the diffraction pattern of the small amount of cobalt, detected under these conditions, is now considered. It seems likely that the cobalt was in one of the three following forms: (*a*) as cobalt sulphate or some other cobalt salt which was not removed during the washing of the specimen after treatment in the plating cell; (*b*) as a uniform metallic film which was in thermodynamic equilibrium with the cobalt ions in the electrolyte; (*c*) as three-dimensional crystallites which were deposited on selected regions of the surface, such as on the peaks of projections.

If the cobalt were in the form of (*a*) or (*b*), it would not give rise to any observable changes in the diffraction patterns. In any case, the observed changes could not have arisen from these types of deposit. Since fairly sharp rings and spots due to twinned crystals are observed, the crystallites giving rise to these patterns must have been of a three-dimensional character. If cobalt had deposited as in case (*c*), then it might be possible for the small amount present to give rise to the observed changes in the diffraction pattern (see Newman and Pashley 1955). If the rings were due to the cobalt, it would have to be concluded that randomly oriented cobalt possessed a pseudomorphic structure. This conclusion is untenable as there can be no basal influence on a randomly oriented polycrystalline aggregate. The spot pattern, however (including the spots due to twinning) could be interpreted as due to pseudomorphic cobalt. If this were so, a local electrolytic cell would be set up on the surface of the copper and cobalt deposition would be expected to continue. (This point is discussed in detail below in connection with the work of Cochrane.) It suffices to give the conclusion that the fixed amount of radioactive cobalt found, independent of the time of immersion in the cell, precludes the interpretation of the spot pattern as due to pseudomorphic cobalt. In support of this conclusion, it may be noted that the intensity of the spots due to twinned crystals increased to a marked extent with increasing periods of immersion of the specimen in the cell at current densities of 40 μA cm^{-2} and 80 μA cm^{-2} while the amount of cobalt present on the surface remained constant. It must therefore be concluded that the deposited cobalt was in form (*a*) or (*b*) and that the new diffraction features were still due entirely to copper.

This latter result can be explained quite simply by a process which is known to occur at the surface of electrodes (see Hillson 1954). During immersion small quantities of copper dissolve from the specimen and are subsequently redeposited. Some of this redeposited copper is twinned with respect to the original crystal and some is randomly oriented. It is impossible to assess the extent to which this process occurs here, since much of the redeposited copper is probably oriented parallel to the bulk crystal and hence does not produce any new diffraction features. This process also accounts for the changes observed in the patterns from specimens which had been immersed in the electrolyte at zero current density. Hence at very low current densities (less than 90 μA cm^{-2})

the question of detecting a pseudomorphic layer does not arise, since no cobalt deposits are formed which are observable by means of electron diffraction.

The explanation of the results obtained at current densities less than $100\,\mu\text{A}\,\text{cm}^{-2}$ with the nickel plating cell is considered to be similar to that given above, although no thickness measurements were possible in this case.

At current densities only slightly greater than the critical values, both nickel and cobalt were deposited as oriented layers. These deposits gave rise to separate spot patterns from the copper substrate and measurements showed that they always had their normal lattice spacings. Some randomly oriented metal was also deposited and the diffraction rings from this passed directly through the corresponding diffraction spots from the oriented metal. Hence it is seen that none of these deposits formed a pseudomorphic layer.

The results obtained by Cochrane (1936) are now discussed in relation to the present work. He did not deposit cobalt at current densities just greater than the critical value, but only at very high and very low values, i.e. $50\,\text{mA}\,\text{cm}^{-2}$ and $4\,\mu\text{A}\,\text{cm}^{-2}$ respectively. At $50\,\text{mA}\,\text{cm}^{-2}$ he obtained only randomly oriented cobalt. He did, however, deposit nickel at a current density of $300\,\mu\text{A}\,\text{cm}^{-2}$ which corresponds to a value only three times greater than the critical value found in the present work when a cleaned anode was used in the plating cell. He obtained a cross-grating pattern of spots with rings, due to a face-centred cubic substance, which passed through the spots. He concluded that the spots were due to copper and the rings were due to randomly oriented nickel. This would lead to the conclusion that the randomly oriented nickel had formed a pseudomorphic layer. It seems likely that in Cochrane's work the pattern from the copper substrate was completely obscured by the nickel deposit. His pattern is thus to be reinterpreted as due to nickel oriented parallel to the copper lattice and to randomly oriented nickel. Hence, no assumption of an anomalous lattice spacing is necessary.

The interpretation of Cochrane's results at low current densities, i.e. less than the critical values, is not clear since no special precautions were taken in his work to obtain repeatable conditions in the plating cell. It has been shown in the present work that if the anode of the cell is not thoroughly cleaned, there is no definite relationship between the applied cell potential and the current density. Whether or not metal is deposited is determined by the applied cell potential and the various overpotentials set up. Hence it is extremely difficult to say whether Cochrane obtained metal deposition in all cases under these conditions. Some of his specimens appeared somewhat darkened but the thickness of the deposits was obviously far less than that calculated from Faraday's law (Cochrane, private communication). This agrees with the results obtained in the present experiments with the nickel cell in which an uncleaned anode was used. In these experiments some nickel deposition definitely occurred and the measured lattice spacing from all specimens was found to be that of copper. This result therefore confirms the findings of Cochrane, although his interpretation is considered to be open to doubt, and an alternative will be put forward.

The darkened appearance of the surface of these specimens was due to the presence of a large number of black specks, as determined by examination with an optical microscope. Thus the nickel had deposited only on to selected regions of the surface. Under these conditions a local electrolytic cell would be set up between these regions and the clean copper surface where no nickel had deposited. As a result, copper would dissolve into the electrolyte and be redeposited on to the nickel. Thus, the surface of the specimen would consist of a thin layer of copper or an alloy of copper and nickel, the main constituent of which would be copper. Lattice spacing measurements would therefore be expected to give a value very close to that of copper.

In the process outlined here, it is assumed that the nickel is not deposited continuously. This assumption is not considered unreasonable since the deposition does not proceed in the usual way (the thickness of the deposit is much less than that expected from Faraday's law, and the nickel is not deposited uniformly over the surface).

An additional reason for rejecting pseudomorphism as an interpretation of these results is that despite a very careful search for the effect in experiments using clean anodes, no evidence for it was in fact found. These experiments were carried out both in the region where deposition starts, the most favourable case, and at higher current densities. In all cases the deposit showed the normal spacing of either cobalt or nickel in the thinnest observable films.

ACKNOWLEDGMENTS

The author is much indebted to Dr. M. Blackman, under whose supervision the work was carried out; to Dr. D. W. Pashley for interest and useful discussions; and to Mr. D. A. Aborav for the provision of the copper single crystals. The author is also indebted to the Department of Scientific and Industrial Research for a maintenance grant.

REFERENCES

BOSWELL, F. W. C., 1950, *Phys. Rev.*, **80,** 91.
COCHRANE, W., 1936, *Proc. Phys. Soc.*, **48,** 723.
FINCH, G. I., and QUARRELL, A. G., 1933, *Proc. Roy. Soc*, A, **144,** 398.
FINCH, G. I., WILMAN, H., and YANG, L., 1947, *Disc. Faraday Soc.*, No. 1, 144.
FRANK, F. C., and VAN DER MERWE, J., 1949, *Proc. Roy. Soc.* A, **198,** 205.
GLASSTONE, S., 1945, *The Electrochemistry of Solutions*, 3rd Edn, (London: Methuen), p. 417.
HILLSON, P. J., 1954, *Trans. Faraday Soc.*, **4,** 385.
KIRCHNER, F., and LASSEN, H., 1935, *Ann. Phys., Lpz.*, **24,** 113.
LUCAS, L. N. D., 1951, *Proc. Phys. Soc.* A, **64,** 943.
NEWMAN, R. C., and PASHLEY, D. W., 1955, *Phil. Mag.*, **46,** 927.
RAETHER, H., 1950, *J. Phys. Radium*, **11,** 11.
SHISHAKOV, N. A., 1952, *Zh. Eksper. Teor. Fiz.*, **22,** 241.
THOMSON, G. P., 1931, *Proc. Roy. Soc.* A, **133,** 1.

The Electrical Resistivity of Sodium between 78°K and 372°K

By F. J. BRADSHAW and S. PEARSON
Royal Aircraft Establishment, Farnborough, Hants.

MS. received 10th October 1955

Abstract. The resistivity was measured to slightly beyond the melting point and results compared with those from previous determinations. There was a small inflection in the curve of (resistivity/temperature) against temperature at about 250°K which could not be simply explained on present conductivity theories and the hypothesis that part of the high temperature resistance is due to lattice defects is considered. The measurements on solid sodium were taken to within 0·04° of the melting point and no further anomalies were found.

Attempts were made to measure excess resistivity due to defects retained after quenching the solid from temperatures near the melting point or after mechanical deformation at 63°K, but any defects so formed proved too mobile for their retention to be possible.

§ 1. Introduction

At temperatures near to their melting points some alkali metals display features in the temperature dependence of their physical properties which on the simple theories would be called anomalous. For example the experimentally determined specific heats at constant pressure of sodium and potassium rise unexpectedly just below their melting points (Carpenter and Steward 1939, Ginnings, Douglas and Ball 1950, Dauphinee, MacDonald and Preston-Thomas 1954). Calculations from these measurements, of the specific heat of the lattice alone at constant volume, based on any reasonable values of the coefficient of expansion and the compressibility, are in excess of $3\boldsymbol{R}$/mole.

The two likely possible explanations are (i) that the thermal vibrations are no longer harmonic and hence the energy of an oscillator per degree of freedom can be greater than $\frac{1}{2}\boldsymbol{k}T$, or (ii) that there are defects formed in the crystal which can contribute independently to the specific heat, resistivity and thermal expansion (MacDonald 1953, Carpenter 1953). Regarding (i), the effects of anharmonic terms are not clear to us except that the observations would demand more than a small non-linear force term to explain them, since this only predicts a linear rise in specific heat with temperature at high temperatures. Regarding (ii) we may be more definite. Localized defects such as Frenkel or Schottky (vacancy) defects must exist, their concentration probably following a simple Boltzmann emperature law. In this paper we think in terms of vacancies.

In an attempt to obtain further information on the subject we have measured the electrical resistance of sodium over a range of temperature. In addition we have taken the opportunity of obtaining absolute resistivity measurements as accurately as possible, because of the scarcity of recent figures.

We have also attempted to produce, by quenching from high temperatures, increases in resistance which might be attributed to defects 'frozen in'. Similarly we have mechanically deformed sodium at low temperatures in order to see whether defects could be produced and retained by this means.

§ 2. RESISTIVITY

2.1. *Experimental Method*

The sodium was obtained in sealed glass ampoules from the Atomic Energy Research Establishment, Harwell. It had been filtered and distilled in a stainless steel still by the method developed by Horsley (1953). Analyses of samples of similarly prepared sodium showed that the purity was better than that normally obtainable by distillation in glass. The principal impurities were potassium, about 25 parts per million, oxygen, less than 5 parts per million. Distillations in glass were shown by Horsley to lead to boron and silicon concentrations in the sodium of about 2000 parts per million.

Nickel tubes of $\frac{1}{2}$ mm diameter, $\frac{1}{40}$ mm wall thickness, 16 mm long were used to contain the sodium. One end of the tube was pressed flat over a length of about 1 mm and then soldered at the tip. This method of sealing prevented any contact between sodium and solder. The tubes were then mounted in a vacuum apparatus, heated to 120°C and evacuated to 10^{-5} mm Hg at this temperature. Liquid sodium was melted out of the ampoule, under vacuum, and driven into the nickel tubes under a pressure of argon which slowly increased to one atmosphere. The tubes were then cooled in a steadily moving sharp temperature gradient starting from the sealed end with the argon pressure still on. When cold, they were nipped off and the end was pressed flat as before, extruding some sodium, and soldered. We think that this method of filling did not affect the purity of the sodium substantially and the method of cooling, in aiming to produce a plane solidification front, prevented the formation of cavities.

Before each tube had been filled two fine copper–constantan thermocouples had been soldered on the outside 10 cm apart, with a third midway between them. These now provided for temperature measurement and also, by using the copper leads only, potential and hence resistance measurements when a current was passed through the specimen. The centre copper lead was used to check that the resistances of the two halves of the specimen were always in the same ratio and hence that the existence of sodium cavities was virtually precluded.

The thermocouples were calibrated against a standard platinum resistance thermometer and the mean temperature from all three thermocouples was always taken. Temperature measurement was accurate to $\frac{1}{40}$ degree over most of the range and to $\frac{1}{20}$ degree below 100°K. The maximum permitted temperature gradient was such that the potential difference between thermocouples was not greater than 2μV. In this paper absolute zero is assumed to be $-273{\cdot}15$°C. Above room temperature the specimens were suspended in a Dewar flask full of oil, with a small electric heater to balance heat loss. Below room temperature, a partially evacuated Dewar filled with an isopentane mixture and surrounded with liquid nitrogen was substituted. Below 115°K baths of liquid oxygen or nitrogen were used.

From x-ray transmission photographs it was estimated that the specimens normally consisted of sodium crystals about 1 cm long occupying the full diameter of the tube. The tubes were cleaned out and remeasured empty to determine the nickel resistance (~10 × sodium resistance) and hence that of the sodium alone. The distance between potential leads was measured by travelling microscope and the tube internal area determined from the length (measured by radiograph) of a weighed amount of mercury. The resistivity of the sodium was hence arrived at, allowance being made for the expansion of the nickel tube with temperature.

2.2. *Results*

Two tubes of sodium were measured and these had resistivities differing by less than $0{\cdot}015 \times 10^{-6}$ ohm cm over the whole temperature range. The mean values are given in the table.

Sodium Resistivity

Temp. (°K)	Res. × 10^6 (ohm cm)	Temp. (°K)	Res. × 10^6 (ohm cm)
70	0·676†	230	3·550
80	0·854	240	3·741
90	1·033	250	3·934
100	1·211	260	4·132
110	1·389	270	4·333
120	1·567	273·15	4·396
130	1·743	280	4·535
140	1·919	290	4·739
150	2·095	300	4·945
160	2·271	310	5·157
170	2·450	320	5·374
180	2·629	330	5·598
190	2·808	340	5·830
200	2·988	350	6·070
210	3·174	360	6·319
220	3·361	370	6·571

† extrapolated from 78°K.

Since the serious errors are likely to be unsuspected systematic ones, it is difficult to give a realistic value for the experimental error. However, considering the likely sources and comparing different independent measurements we estimate the average differential error in resistivity to be $\pm 0{\cdot}006 \times 10^{-6}$ ohm cm (and less than this for small temperature intervals). For absolute values the figure might be $\pm 0{\cdot}01 \times 10^{-6}$ ohm cm.

§ 3. QUENCHING AND COLD WORK EXPERIMENTS

An attempt was made to see if any part of the resistance at high temperature could be retained on quenching and possibly be attributed to lattice defects. It was considered more important to have a rapid initial cooling rather than a very low end temperature if any defects were to be retained, since their mobility might well be expected to vary exponentially with temperature. Independent experiments had shown that it was better to quench into cooled organic liquids with a wide liquid temperature range rather than into boiling liquid gases such as nitrogen. Accordingly the isopentane mixture at 117°K was used as a quenching medium. Specimens were heated to within one degree of the sodium melting point and were then dropped into the mixture. The rates of cooling under these conditions were such that the time to reach half the total temperature interval (~240°K) was about 50 milliseconds. However, it was found that the change in resistivity of the specimens after heating and quenching was always less than $0{\cdot}005 \times 10^{-6}$ ohm cm and therefore insignificant.

Further annealed nickel tubes 10 cm long with two thermocouples at 5 cm apart were soldered into threaded bushes and filled with sodium as before. The bushes were screwed into a simple rig made to extend a specimen at low temperatures. This was placed in liquid nitrogen in a copper can, the whole being surrounded with liquid oxygen boiling under reduced pressure at 63°K.

The specimen was then extended 10% in length, and the change of resistance measured as quickly as possible, the measurement being completed approximately two minutes after extension. Subsequent measurements up to one hour later showed no change of resistance with time. The rig and specimen were then removed into liquid nitrogen at 78°K (during transfer the temperature did not rise above 78°K) and the resistance remeasured. Again no change with time was observed. The specimen was heated to room temperature and ultimately to within eight degrees of the melting point and then measured in liquid nitrogen. Again no change in resistivity occurred. We conclude from this and a similar repeat experiment that either there was an increase in resistivity which did not anneal out within an hour at 63°K and annealed out in less than two minutes at 78°K or that any increase in resistivity due to deformation could not be retained even at 63°K. The first alternative is very unlikely and so we say that the increase in resistance of the specimen on pulling at 63°K was due solely to the change in specimen dimensions.

§ 4. Resistivity near the Melting Point

To check that no anomalies were occurring within the last degree or so before melting, measurements of resistivity were made as close to the melting point as possible. Arrangements were made for the oil bath temperature to rise or fall very slowly, i.e. at rates of $\frac{1}{2}$° to 1° per hour. Resistance measurements on one specimen were taken at temperature intervals $\frac{1}{10}$ to $\frac{1}{40}$ degree with the temperature slowly increasing, until the resistance started suddenly to increase rapidly. This point was taken to be the onset of melting (see figure 1, Run 1). As soon as the resistance reached a steady value again, further measurements were taken, and this time the sodium was assumed to be all liquid. After raising the temperature to about one degree above the melting point, the specimen was cooled similarly; it was found possible to supercool the sodium by about $\frac{1}{20}$ degree. As might be expected on solidifying, the resistivity did not recover to the original solid value since cavities formed. But by re-melting the sodium and cooling it with a moving temperature gradient as during the filling, it was possible to repeat the original solid measurements almost exactly, i.e. the cavities were eliminated (Run 2). As can be seen the measurements on the liquid sodium in Run 2 did not repeat as well as previously. This, we suggest, may be due to a slight expansion of the nickel tube due to the sodium expanding on melting for the second time. However, since the solid measurements in Run 2 repeated closely those in Run 1, there can have been effectively no tube expansion in the first melting. For this reason we have taken Run 1, solid and liquid, and Run 2, solid only, as giving the most reliable figures.

The results in figure 1 showed clearly that the melting point lay between 370·92° and 371·00°K the last steady solid and first steady liquid readings respectively. If we take the mean value of 370·96°K to be the melting temperature then the resistivities of solid and liquid sodium at this temperature are $6{\cdot}598 \times 10^{-6}$ and $9{\cdot}573 \times 10^{-6}$ ohm cm respectively, giving a ratio of 1 : 1·451. The slopes of the solid and liquid lines at the melting point were found to be $0{\cdot}0265 \times 10^{-6}$ and $0{\cdot}0347 \times 10^{-6}$ ohm cm deg^{-1}.

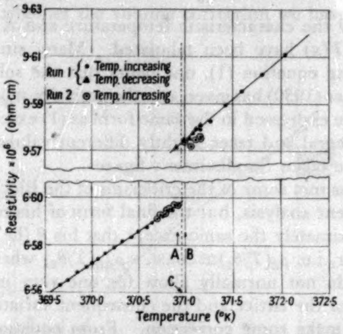


Figure 1. Resistivity of sodium near the melting point. A, last steady solid reading, temperature increasing ; B, first steady liquid reading, temperature increasing ; C, melting point (Ginnings *et al.* 1950).

§ 5. DISCUSSION

5.1. *Comparison with Previous Work*

We have found no recently published absolute determinations of resistivity. Hackspill's (1910) figure of $4{\cdot}27 \times 10^{-6}$ ohm cm at 0°c often appears to be accepted though he seems to have selected results from experiments giving low values of resistivity, rejecting higher ones on the grounds that they were due to impurities, etc. At 100°c we obtain by extrapolation a resistivity of the liquid of $9{\cdot}649 \times 10^{-6}$ ohm cm; this agrees well with Lyon's (1952) figure of $9{\cdot}65 \times 10^{-6}$ ohm cm.

More has been published on relative values of resistivity. MacDonald (1952) has summarized some of the earlier determinations. The values of (resistivity)/(resistivity at 0°c) which he gives are from 0 to 3% different from ours. There are differences similarly between MacDonald's (1953) recent data and ours of up to 5%. We do not observe his sudden rise in the resistivity temperature curve near to the melting point, but simply have a gentle curvature starting from about 200°K. In this respect we agree more with Bidwell's (1924) results, except that very near to the melting point he obtained a large resistivity rise (perhaps due to impurities).

The melting temperature of sodium at one atmosphere has been given by Ginnings *et al.* (1950) as being 370·96°K. This agrees well with our own figure. The sharpness of melting or the fact that no anomaly appeared in the resisitivity up to within about 0·04° of the melting point is a good indication of the purity of the specimen.

5.2. *Comparison with Theory*

The Bloch–Grüneisen formula for resistivity ρ of a metal is

$$\frac{\rho}{T}=\frac{K}{\theta^2}F\left(\frac{\theta}{T}\right) \qquad \ldots\ldots(1)$$

where

$$F(x)=x^{-4}\int_0^x z^5/(e^z-1)(1-e^{-z})dz,$$

T is temperature, θ the characteristic temperature and K is a constant for the metal. Values of $F(x)$ have been tabulated. Many simplifying assumptions are made in deriving equation (1), one being that the solid follows the Debye model. Sondheimer (1950) has given a mathematically more rigorous derivation for ρ/T which can be expressed in the same form as (1) except that $F(x)$ is now no longer a simple integral and takes slightly different values. In the subsequent calculations we have taken Sondheimer's figures.

Ziman (1954) has met some of the criticisms of the Bloch–Grüneisen assumptions in a more recent analysis, but the final form of his resistance–temperature curve is still approximately the same except that his θ (he assumes a Debye-like solid) is now higher, i.e. $\rho_Z(T/\theta_1)\simeq \text{const.}\times\rho_{BG}(T/\theta_2)$ where $\theta_2\simeq 0{\cdot}86\,\theta_1$.

The formulae do not normally allow for one very important feature, the thermal expansion of the lattice and the consequent variation of K and θ. It is essential for us to make some correction. From equation (1) we may obtain an expression for the correction $\Delta\rho/\rho$ due to a change in volume in expanding from $T=0$ in terms of $\Delta K/K$ and $\Delta\theta/\theta$. From Grüneisen's law we can calculate the contribution due to $\Delta\theta/\theta$ in terms of $\Delta V/V$ at a given temperature (we assume for sodium $\gamma=1{\cdot}25$). Assume that $\Delta K/K$ depends only on V, i.e.

$$\Delta K/K=A(\Delta V/V)+B(\Delta V/V)^2 \qquad \ldots\ldots(2)$$

and we have $\Delta\rho/\rho$ in terms of $\Delta V/V$ for any value of T with only the constants A and B undetermined. Bridgman's (1949) data on the change of resistance with volume at one temperature enables us to determine these constants. Taking what we understand to be his best figures, we have

$$\Delta\rho/\rho=4{\cdot}85(\Delta V/V)-13{\cdot}9(\Delta V/V)^2 \qquad \ldots\ldots(3)$$

as being the relation at 20°C.

Calculations show that the effect of varying temperature is small, i.e. the constants in (3) only change appreciably when $\Delta V/V$ is small so we may regard (3) as applying to all temperatures. For the dependence of $\Delta V/V$ on temperature we have used the experiments of Siegel and Quimby (1938) and Griffiths and Griffiths (1915).

Volume corrections could of course be made using only the first power of $\Delta V/V$. Kelly (1954) has done this. But we have found that not only has the $(\Delta V/V)^2$ term a substantial effect ($\Delta V/V$ from 0°K to the melting point is $\sim$6%) but also that by ignoring it, a plot of experimental ρ/T corrected to constant volume against T has a region of negative slope at about 300°K. Such a feature would be unexpected here.

In figure 2 the experimental $(\rho/T, T)$ is compared with Sondheimer's theoretical curves corrected for thermal expansion. The latter are drawn for different values of θ and are fitted, by allowing K to vary arbitrarily, to the experimental curve at points where they have common slopes. It can be seen that for any one value of θ the fit is limited, giving approximately

T(°K)	150	200	250
θ(°K)	160	120	50
K/θ^2(10^{-8} ohm cm deg^{-1})	1·39	1·36	1·35

(Using Ziman's calculations θ would have been ~16% higher.)

The above rapid fall in θ with increasing T makes comparison with θ values derived from other physical properties, such as specific heat, pointless. It also raises questions as to whether the volume correction we have taken may not be too much.

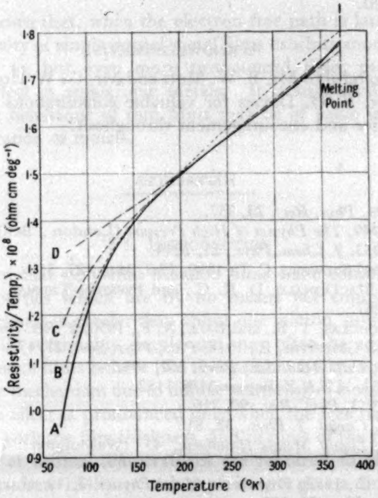


Figure 2. Restivity of sodium ; experimental and theoretical curves. A, Experimental; B, theoretical, $\theta=160$°K, fitted at 150°K; C, theoretical, $\theta=120$°K, fitted at 200°K ; D, theoretical $\theta=50$°K, fitted at 250°K.

However, it is clear that no simple extension of the present theories will account for the point of inflection in the (ρ/T, T) curve in the region of 250°K. This could be attributed to the presence of defects, though without any direct justification. In this respect we agree with MacDonald (1953) but differ from him in speculating on the number involved. Jongenburger (1953) has calculated the effect of vacancies on the resistivity of copper. If we assume a similar order of magnitude for their effect in sodium, then the difference between experimental and the $\theta=50$°K theoretical curve at the melting point is due to a $\Delta\rho$ of about 7×10^{-8} ohm cm which in turn would be caused by about 7×10^{-4} concentration of vacancies. Assuming that the concentration is $A\exp(-F/\boldsymbol{k}T)$, where $A\sim1$, this implies an energy of formation of 0·2 ev which may be compared with the energy of 0·45 ev for self-diffusion in sodium (Nachtrieb, Catalano and Weil 1952). The difference of 0·25 ev would have to be considered as the energy for vacancy movement.

With regard to the cold work experiments, it would be expected from Manintveld's (1952) work that if the defects and dislocations which must have been produced in the sodium by cold work had been retained at 63°K then the 10% extension would have lead to an annealable change in resistivity appreciably greater than the limits of measurement. In other words, any defects formed during the extension were mobile enough to cluster or vanish even at 63°K. An energy of 0·25 ev for movement would not be unreasonable for this to happen. It also implies that the strain which the sodium must have undergone in conforming to the size of the nickel tube without cavities would have no effect on resistivity, since at all temperatures such strains were able to anneal sufficiently to leave the resistivity unaltered.

ACKNOWLEDGMENTS

We are most grateful to Mr. B. W. Mott and Mr. G. W. Horsley for supplying the sodium, to Dr. R. O. Davies for valuable consultations and to Mr. L. G. Carpenter for advice and encouragement throughout.

REFERENCES

BIDWELL, C. G., 1924, *Phys. Rev.*, **23,** 357.
BRIDGMAN, P. W., 1949, *The Physics of High Pressure* (London : Bell).
CARPENTER, L. G. 1953, *J. Chem. Phys.*, **21,** 2244.
CARPENTER, L. G., and STEWARD, C. J., 1939, *Phil. Mag.*, **27,** 551.
DAUPHINEE, T. M., MACDONALD, D. K. C., and PRESTON-THOMAS, H., 1954, *Proc Roy. Soc.* A, **221,** 267.
GINNINGS, D. C., DOUGLAS, T. B., and BALL, A. F., 1950, *J. Res. Nat. Bur. Stds.*, **45,** 23.
GRIFFITHS, E. A., and GRIFFITHS, E., 1915, *Proc. Phys. Soc.*, **27,** 477.
HACKSPILL, L., 1910, *C.R. Acad. Sci., Paris*, **151,** 305.
HORSLEY, G. W., 1953, *A.E.R.E.* Report M/R 1152.
JONGENBURGER, P., 1953, *Phys. Rev.*, **90,** 710.
KELLY, F. M., 1954, *Canad. J. Phys.*, **32,** 81.
LYON, R. M., 1952, *Liquid Metals Handbook*, 43. (Washington : U.S. Atomic Energy Commission).
MACDONALD, D. K. C., 1952, *Progress in Metal Physics*, **3,** 42. (London : Butterworth); 1953, *J. Chem. Phys*, **21,** 177.
MANINTVELD, J. A., 1952, *Nature, Lond.*, **169,** 623.
NACHTRIEB, N. H., CATALANO, E., and WEIL, J. A., 1952, *J. Chem. Phys.*, **20,** 1185.
SIEGEL, S., and QUIMBY, S. L., 1938, *Phys, Rev.*, **54,** 76.
SONDHEIMER, E. H., 1950, *Proc. Roy. Soc.* A, **203,** 75.
ZIMAN, J. M., 1954, *Proc. Roy. Soc.* A, **226,** 436.

The Electrical Conductivity of Anisotropic Thin Films

By R. ENGLMAN and E. H. SONDHEIMER
Department of Mathematics, Queen Mary College, London

MS. received 5th December 1955

Abstract. It is shown that, when the electron free path is large, the theoretical electrical conductivity of single crystal metal films exhibits anomalous anisotropic properties similar to, but even more pronounced than, those found in the anomalous skin effect in anisotropic metals. It is suggested that experiments on the anisotropic behaviour of thin films may be of value in the study of the electronic configuration of metals.

§ 1. Introduction

It is well known that very thin metallic films have complicated and varied electrical properties which are by no means yet completely understood. If the film thickness is greater than about one micron, however, the electrical behaviour is usually reasonably simple and may often be adequately described by assuming that such films behave like the metal in bulk, except that there is an additional resistive mechanism due to diffuse scattering of electrons at the surfaces of the film. This effect is pronounced only when the free path of the electrons is large compared with the film thickness, and in films of the thickness envisaged here experiments at helium temperatures are required to obtain the requisite long free paths (Andrew 1949). Such experiments give a direct estimate of the free path of electrons in metals. A related type of phenomenon is the so-called anomalous skin effect in metals, where the free path is compared with the depth of penetration of a high-frequency electromagnetic field into the metal (Chambers 1952).

Recent experimental and theoretical work on the anomalous skin effect (Pippard 1950, 1954, Sondheimer 1954, to be referred to as S) has shown that these mean free path effects provide particularly valuable information about the electronic structure of metals when the experiments are carried out on single crystals. Ordinarily, the electrical conductivity σ of an anisotropic metal is represented by a symmetric second-order tensor, and the anisotropy of σ is determined by the direction cosines specifying the direction of current flow relative to the crystal axes. The situation in thin films and in the anomalous skin effect is more complicated since, when surface effects are appreciable, electrons moving at different angles to the surface of the metal contribute in different amounts to the total current. For different orientations of the crystal axes relative to the surface, therefore, different groups of electrons become effective, and the resistance depends not only on the direction of current flow but also on the orientation of the surface relative to the crystal axes. In this way the anisotropy of the resistance reflects the local geometry of the Fermi surface of the metal.

The theory of the anomalous skin effect has been developed in S for the simplest model of a uniaxial anisotropic metal, in which the surfaces of constant energy are assumed to be families of ellipsoids of revolution about the principal axis of the crystal. In the present paper we derive the corresponding expressions for the conductivity of a thin metallic anisotropic film subjected to a d.c. electric field (the analogous problem for semiconductors has recently been considered by Ham and Mattis (1955)). The model chosen, although somewhat artificial, has the advantage that the calculations can be carried out exactly, and in fact in the present problem it is possible to go further than in S and to derive explicit expressions (equations (14) and (16)) for the conductivity valid over the whole range of values of the free path. The results show that the anisotropy of the d.c. resistance of thin films is analogous to, but even more pronounced than, the anisotropy of the surface resistance in the anomalous skin effect. To our knowledge there are at present no experimental results with which to compare the predictions of the present theory, and in view of the artificial nature of the model the results obtained here have mainly illustrative significance. The results suggest, however, that experiments on the anisotropy of the resistance of single crystal metal films, if practicable, should be of interest as a tool in the study of the electronic configuration of metals.

§ 2. The Current Densities

2.1.

We consider a single crystal metal film bounded by the planes $z=0$ and $z=b$. Only crystals with a single principal axis will be considered, and we suppose that the crystal axis lies in the plane $y=0$ and makes an angle ψ with the z axis. In the limiting case of very thick films, when effects due to the surface of the film are negligible, the relation between the components J_i $(i=1, 2, 3)$ of the electric current and the (constant) electric field components $\mathscr{E}_i$ is

$$J_i = \sum \sigma_{ij} \mathscr{E}_j \qquad \ldots\ldots(1)$$

where the conductivity tensor σ_{ij} may be written in the form

$$\sigma \begin{bmatrix} 1+a\sin^2\psi & 0 & a\sin\psi\cos\psi \\ 0 & 1 & 0 \\ a\sin\psi\cos\psi & 0 & 1+a\cos^2\psi \end{bmatrix} \qquad \ldots\ldots(2)$$

σ and $\sigma(1+a)$ being the conductivities for current flow perpendicular and parallel to the principal axis of the crystal. For comparison with experiment we require the conductivities σ_1 and σ_2 in the two principal directions in the plane of the film, i.e. for current flow in and perpendicular to the plane $y=0$. These are given by $J_1/\mathscr{E}_1$ and $J_2/\mathscr{E}_2$, where $J_1/\mathscr{E}_1$ is to be evaluated by imposing the condition that the transverse current is zero $(J_3=0)$ and eliminating $\mathscr{E}_3$ between the expressions for J_1 and J_3 obtained from (1) and (2). This gives

$$\sigma_1 = \sigma \frac{1+a}{1+a\cos^2\psi}, \qquad \sigma_2 = \sigma. \qquad \ldots\ldots(3)$$

When σ_1 and σ_2 are known, the conductivity $\sigma(\phi)$ for current flowing at an angle ϕ to the x axis can be obtained from the relation

$$\frac{1}{\sigma(\phi)} = \frac{\cos^2\phi}{\sigma_1} + \frac{\sin^2\phi}{\sigma_2}.$$

2.2.

In the general case of a film of arbitrary thickness it is necessary to calculate the generalized relation between $\mathbf{J}$ and $\mathfrak{E}$ which replaces (1) and (2) when surface effects are taken into account. In order to carry out the calculations explicitly special models must be considered; we shall adopt the simplest and most convenient in which the electronic energy surfaces are assumed to form families of ellipsoids of revolution about the principal axis of the crystal, and in which the electronic time of relaxation is a function of the energy only. For this model the required relation between $\mathbf{J}$ and $\mathfrak{E}$ has, in effect, already been derived in S in discussing the anomalous skin effect in anisotropic metals, for the case (the most important one in practice and the only one to be considered here) in which the electrons suffer diffuse reflection at the surface of the metal.

However, the present case differs from that treated in S in two aspects. First, a film of finite thickness b is considered here instead of a semi-infinite slab; this means (as is easily verified) that the range of integration $(0, \infty)$ in all integrals is to be replaced by the range $(0, b)$. Secondly, in the high frequency case the electric field components must all be treated as unknown functions of z, whereas in the present case we have curl $\mathfrak{E}=0$, so that $\mathscr{E}_1$ and $\mathscr{E}_2$ are constant and only the transverse field $\mathscr{E}_3$ varies with z.

The correct expressions for the current may now be written down by making the required modifications in Sondheimer's expressions. As in S, we consider for generality the case in which the conduction electrons occupy two energy bands, but neglect transitions between the bands, and the currents are then given by (compare S equations (30) to (32), (35) to (38))†

$$\begin{aligned} J_1(z) = \frac{3}{4}\mathscr{E}_1 \int_0^b \Big\{ & \frac{\sigma'}{L'}\frac{1+a'}{1+a'\cos^2\psi}(\mathrm{Ei}_1' - \mathrm{Ei}_3') + \frac{\sigma''}{L''}\frac{1+a''}{1+a''\cos^2\psi}(\mathrm{Ei}_1'' - \mathrm{Ei}_3'') \\ & + 2\sin^2\psi\cos^2\psi\left(\frac{\sigma'}{L'}\frac{a'^2}{1+a'\cos^2\psi}\mathrm{Ei}_3' + \frac{\sigma''}{L''}\frac{a''^2}{1+a''\cos^2\psi}\mathrm{Ei}_3''\right)\Big\}\,dt \\ & + \frac{3}{2}\sin\psi\cos\psi\int_0^b\left(\frac{\sigma'a'}{L'}\mathrm{Ei}_3' + \frac{\sigma''a''}{L''}\mathrm{Ei}_3''\right)\mathscr{E}_3(t)\,dt \qquad \ldots\ldots(4) \end{aligned}$$

$$J_2(z) = \frac{3}{4}\mathscr{E}_2\int_0^b\left\{\frac{\sigma'}{L'}(\mathrm{Ei}_1' - \mathrm{Ei}_3') + \frac{\sigma''}{L''}(\mathrm{Ei}_1'' - \mathrm{Ei}_3'')\right\}dt \qquad \ldots\ldots(5)$$

$$\begin{aligned} J_3(z) = {} & \frac{3}{2}\sin\psi\cos\psi\,\mathscr{E}_1\int_0^b\left(\frac{\sigma'a'}{L'}\mathrm{Ei}_3' + \frac{\sigma''a''}{L''}\mathrm{Ei}_3''\right)dt \\ & + \frac{3}{2}\int_0^b\left\{\frac{\sigma'}{L'}(1+a'\cos^2\psi)\mathrm{Ei}_3' + \frac{\sigma''}{L''}(1+a''\cos^2\psi)\mathrm{Ei}_3''\right\}\mathscr{E}_3(t)\,dt \qquad \ldots\ldots(6) \end{aligned}$$

where σ', $\sigma'(1+a')$ are the bulk metal conductivities perpendicular and parallel to the crystal axis for electrons in the first band only,

$$L' = l'(1+a'\cos^2\psi)^{1/2} \qquad \ldots\ldots(7)$$

---

† These expressions, like those in S, are valid subject to the assumption that the form of the collision term in the Boltzmann equation for the electronic distribution function is the same as when surface effects are absent. This assumption has been discussed by Ham and Mattis (1955) who conclude that, although the assumption cannot be strictly valid, the errors introduced by it are small (see also MacDonald and Sarginson 1950).

l' being a constant of the order of the free path for electrons in the first band, and

$$\mathrm{Ei}_n' = \mathrm{Ei}_n(|z-t|/L') = \int_1^\infty s^{-n}\exp(-s|z-t|/L')ds; \qquad \ldots\ldots(8)$$

the corresponding quantities for electrons in the second band are defined similarly.

§ 3. THE CONDUCTIVITIES

3.1.

The principal conductivity σ_1 for current flow in the plane $y=0$ is defined by

$$\bar{J}_1(z)/\mathscr{E}_1 = b^{-1}\int_0^b J_1(z)dz/\mathscr{E}_1,$$

and is to be evaluated by eliminating $\mathscr{E}_3$ between equations (4) and (6), putting $J_3(z)=0$ in (6) (since curl $\mathbf{H}$ has zero z component). The elimination of $\mathscr{E}_3$ is elementary in the particular case when only one energy band is considered, but the results of S indicate that this case is insufficiently general to give results of physical interest; it is considered briefly in § 3.2.2 below. In the general case of the two-band model the exact solution can be obtained by means of the theory of Fourier transforms.

For this purpose we note that

$$\begin{aligned}\int_{-\infty}^{\infty} e^{iuz}dz \int_0^b \mathrm{Ei}_3(|z-t|/L)\mathscr{E}_3(t)dt &= \int_0^b \mathscr{E}_3(t)dt\int_{-\infty}^{\infty}\mathrm{Ei}_3(|z-t|/L)e^{iuz}dz \\ &= \int_0^b \mathscr{E}_3(t)e^{iut}dt\int_{-\infty}^{\infty}\mathrm{Ei}_3(|w|/L)e^{iuw}dw \\ &= LI(Lu)\mathscr{F}(u), \qquad \ldots\ldots(9)\end{aligned}$$

where

$$\mathscr{F}(u) = \int_0^b \mathscr{E}_3(t)e^{iut}dt \qquad \ldots\ldots(10)$$

and

$$I(u) = \int_{-\infty}^{\infty}\mathrm{Ei}_3(|t|)e^{iut}dt = 2u^{-3}(u-\tan^{-1}u). \qquad \ldots\ldots(11)$$

Inverting [illegible] Fourier transform (9), we see that the second integral in (4) may be written

$$\begin{aligned}\int_0^b\left\{\frac{\sigma'a'}{L'}\mathrm{Ei}_3\left(\frac{|z-t|}{L'}\right)+\frac{\sigma''a''}{L''}\mathrm{Ei}_3\left(\frac{|z-t|}{L''}\right)\right\}\mathscr{E}_3(t)dt \\ = \frac{1}{2\pi}\int_{-\infty}^{\infty}\{\sigma'a'I(L't)+\sigma''a''I(L''t)\}\mathscr{F}(t)e^{-izt}dt. \qquad \ldots\ldots(12)\end{aligned}$$

Also the Fourier transform of (6) is (remember that $\mathscr{E}_1$ is constant)

$$\begin{aligned}&\int_{-\infty}^{\infty} J_3(z)e^{itz}dz \\ &\quad = \tfrac{3}{2}\sin\psi\cos\psi\,\mathscr{E}_1\frac{e^{ibt}-1}{it}\{\sigma'a'I(L't)+\sigma''a''I(L''t)\} \\ &\qquad + \tfrac{3}{2}\{\sigma'(1+a'\cos^2\psi)I(L't)+\sigma''(1+a''\cos^2\psi)I(L''t)\}\mathscr{F}(t) \qquad \ldots\ldots(13) \\ &\quad = 0,\end{aligned}$$

since $J_3(z)=0$. This determines $\mathscr{F}(t)$, and $\mathscr{E}_3(z)$ can be obtained, if required, as a Fourier series by inverting the 'finite transform' (10). Substituting the expression for $\mathscr{F}(t)$ obtained from (13) into (12) and (4), we obtain

$$\begin{aligned} J_1(z)=\frac{3}{4}\,\mathscr{E}_1\int_0^b & \left\{\frac{\sigma'}{L'}\frac{1+a'}{1+a'\cos^2\psi}(\mathrm{Ei}_1'-\mathrm{Ei}_3')+\frac{\sigma''}{L''}\frac{1+a''}{1+a''\cos^2\psi}(\mathrm{Ei}_1''-\mathrm{Ei}_3'')\right. \\ & \left.+2\sin^2\psi\cos^2\psi\left(\frac{\sigma'}{L'}\frac{a'^2}{1+a'\cos^2\psi}\mathrm{Ei}_3'+\frac{\sigma''}{L''}\frac{a''^2}{1+a''\cos^2\psi}\mathrm{Ei}_3''\right)\right\}dt \\ & -\frac{3}{4\pi}\sin^2\psi\cos^2\psi\,\mathscr{E}_1\int_{-\infty}^{\infty}\frac{\{\sigma'a'I(L't)+\sigma''a''I(L''t)\}^2}{\sigma'(1+a'\cos^2\psi)I(L't)+\sigma''(1+a''\cos^2\psi)I(L''t)} \\ & \times\left(\frac{\mathrm{e}^{ibt}-1}{it}\right)\mathrm{e}^{-izt}dt. \end{aligned}$$

This gives, on performing the integration over z,

$$\begin{aligned} \sigma_1=\sigma'\frac{1+a'}{1+a'\cos^2\psi}K_1(b/L')&+\sigma''\frac{1+a''}{1+a''\cos^2\psi}K_1(b/L'') \\ & +\sin^2\psi\cos^2\psi\left\{\frac{\sigma'a'^2}{1+a'\cos^2\psi}K_2(b/L')+\frac{\sigma''a''^2}{1+a''\cos^2\psi}K_2(b/L'')\right\} \\ & -\frac{3b}{2\pi}\sin^2\psi\cos^2\psi\int_0^{\infty}\frac{\{\sigma'a'I(L't)+\sigma''a''I(L''t)\}^2}{\sigma'(1+a'\cos^2\psi)I(L't)+\sigma''(1+a''\cos^2\psi)I(L''t)} \\ & \times\left(\frac{\sin\frac{1}{2}bt}{\frac{1}{2}bt}\right)^2 dt \qquad \ldots\ldots(14) \end{aligned}$$

where

$$\left.\begin{aligned} K_1(u)&=1-\frac{3}{8u}+\frac{3}{2u}\{\mathrm{Ei}_3(u)-\mathrm{Ei}_5(u)\} \\ K_2(u)&=1-\frac{3}{4u}+\frac{3}{u}\mathrm{Ei}_5(u) \end{aligned}\right\} \qquad \ldots\ldots(15)$$

$\mathrm{Ei}_n(u)$ being defined by (8) and $I(u)$ by (11). It should be noted that in the analogous high-frequency problem considered in S it was not found possible to derive an exact solution corresponding to (14).

The evaluation of the conductivity in the second principal direction is much easier, and from (5) we obtain at once

$$\sigma_2=b^{-1}\int_0^b J_2(z)dz/\mathscr{E}_2=\sigma'K_1(b/L')+\sigma''K_1(b/L$$ [illegible] ……(16)

Equations (14) and (16) represent the exact solutions of the present problem and can now be used to discuss the different cases of interest. It is instructive, first, to consider various cases for which simple limiting expressions can be derived.

3.2. *Very Thin Films*

When u is small, $K_1(u)=\frac{3}{4}u\log u^{-1}+\mathrm{O}(u)$ while $K_2(u)$ is $\mathrm{O}(u)$; it may further be shown that, when b/L' and b/L'' are both small compared with unity, the last term in (14) is small of order b/L' (or b/L''). Hence the leading terms in σ_1 and σ_2 for very thin films are

$$\sigma_1=\frac{3}{4}\frac{\sigma'(1+a')}{1+a'\cos^2\psi}\frac{b}{L'}\log\frac{L'}{b}+\frac{3}{4}\frac{\sigma''(1+a'')}{1+a''\cos^2\psi}\frac{b}{L''}\log\frac{L''}{b} \qquad \ldots\ldots(17)$$

$$\sigma_2=\frac{3}{4}\sigma'\frac{b}{L'}\log\frac{L'}{b}+\frac{3}{4}\sigma''\frac{b}{L''}\log\frac{L''}{b}. \qquad \ldots\ldots(18)$$

It should be noted that, in this approximation, σ_1 and σ_2 are both composed additively of contributions from the two energy bands separately, whereas in the general case this is true for σ_2 but not for σ_1; further, σ_1 and σ_2 depend upon the free path (through the logarithmic factors) even when the free path is very long. In the analogous skin effect problem (see S, equations (47) and (45)) $1/Z_1{}^3$ and $1/Z_2{}^3$ are additive for the two bands when the free path is very long, where Z_1 and Z_2 are the surface impedances corresponding to σ_1 and σ_2, and Z_1 and Z_2 are independent of the free path in this limit.

3.3. *Thick Films*

For large values of b/L' and b/L'' the leading terms in σ_1 and σ_2 are obtained by replacing $K_1(u)$ and $K_2(u)$ by their value 1 for large u, and by replacing $I(t)$ in the definite integral in (14) by its value $\frac{2}{3}$ for small t. Then the integral reduces to

$$\frac{2}{3}\frac{(\sigma'a'+\sigma''a'')^2}{\sigma'(1+a'\cos^2\psi)+\sigma''(1+a''\cos^2\psi)}\int_0^\infty\left(\frac{\sin\frac{1}{2}bt}{\frac{1}{2}bt}\right)^2 dt$$
$$=\frac{2\pi}{3b}\frac{(\sigma'a'+\sigma''a'')^2}{\sigma'(1+a'\cos^2\psi)+\sigma''(1+a''\cos^2\psi)}.$$

But
$$\sigma'+\sigma''=\sigma,\quad \sigma'(1+a')+\sigma''(1+a'')=\sigma(1+a) \qquad \ldots\ldots(19)$$
where σ and $\sigma(1+a)$ are the total bulk conductivities defined in § 2.1, and a simple reduction now leads to

$$\sigma_1=\sigma\frac{1+a}{1+a\cos^2\psi},\qquad \sigma_2=\sigma \qquad \ldots\ldots(20)$$

in agreement with (3).

It is possible to derive higher terms in the expansions of (14) for very thin and very thick films, but these will not be given here since they involve complicated definite integrals, so that there is no appreciable gain in simplicity over the general expression (14). In the case of (16) the higher terms in the expansions are readily written down if required, by making use of the well-known properties of the integrals $\mathrm{Ei}_n(u)$ (see Kourganoff 1952, Appendix). (In the case of thick films the integrand of the definite integral in (14) oscillates rapidly. For the purpose of computation it is then best to convert the integral into a form with a non-oscillating integrand, by a generalization of the method used in the Appendix.)

3.4. *Single Band*

When only one band is considered, we may evaluate σ_1 directly by multiplying equation (6) by $a\sin\psi\cos\psi(1+a\cos^2\psi)^{-1}$ (dropping the dashes for convenience) and subtracting from (4). This eliminates $\mathscr{E}_3$, and we obtain

$$\sigma_1=\sigma\frac{1+a}{1+a\cos^2\psi}K_1(b/L). \qquad \ldots\ldots(21)$$

On the other hand, for a single band the definite integral in (14) reduces to

$$\frac{\sigma a^2}{1+a\cos^2\psi}\int_0^\infty I(Lt)\left(\frac{\sin\frac{1}{2}bt}{\frac{1}{2}bt}\right)^2 dt=\frac{2\sigma a^2}{b(1+a\cos^2\psi)}\int_0^\infty I\left(\frac{2Ls}{b}\right)\frac{\sin^2 s}{s^2}\,ds,$$

and therefore, to show that (14) reduces to (21) for a single band, we have to prove that

$$\frac{3}{\pi}\int_0^\infty I\left(\frac{2s}{y}\right)\frac{\sin^2 s}{s^2}\,ds=K_2(y)\qquad (y>0) \qquad \ldots\ldots(22)$$

where $I(u)$ is given by (11) and $K_2(u)$ by (15). The proof is given in the Appendix.

3.5. *Numerical Results*

The properties of the generalized exponential integrals $\mathrm{Ei}_n(u)$ required in the evaluation of σ_1 and σ_2 in the general case are well known; in particular, they can all be expressed in terms of the tabulated integral $\mathrm{Ei}_1(u)$. The definite integral in (14), however, must in general be evaluated by numerical integration.

Chief interest attaches to the anisotropy of σ_1 and σ_2 for different values of the ratio film thickness/free path; this may be conveniently exhibited by plotting the ratios $\sigma_1(\psi)/\sigma_1(\frac{1}{2}\pi)$ and $\sigma_2(\psi)/\sigma_2(\frac{1}{2}\pi)$ as functions of the angle ψ between the crystal axis and the normal to the film surface. Using the relations (19) it is easily shown that, for a given value of the parameter a (which is determined by the observed anisotropy of the conductivity of the metal in bulk), the ratios $\sigma_1(\psi)/\sigma_1(\frac{1}{2}\pi)$ and $\sigma_2(\psi)/\sigma_2(\frac{1}{2}\pi)$ for the two-band model involve four independent parameters (compare the discussion in S, where, however, only the limit of very long free paths is considered); these four may be taken to be the anisotropy parameters a' and a'' for the two bands separately, the ratio l'/l'' of the free paths in the two bands (we recall that L', L'' are related to l', l'' by (7)), and the ratio b/l'' which is a measure of the thickness of the film relative to the electronic free path. To illustrate the general nature of the results we have plotted in figures 1 and 2 the curves obtained for various values of the thickness parameter b/l'', taking $a=-0{\cdot}12$, $a'=5$, $a''=-0{\cdot}95$, $l'/l''=0{\cdot}41$; these values of the parameters were chosen, in the absence of direct experimental data on the effect discussed here, because they were found to fit the experimental results of Pippard (1950) on the anomalous skin effect in single crystals of tin (see S. Recent experimental work by Fawcett (1955) has shown that the simple model considered in the present paper is, in fact, not sufficiently general to represent the behaviour of tin, since the observed high-frequency resistance is markedly anisotropic in the plane perpendicular to the tetrad axis of the crystal. While the present theory could be generalized so as to take more complicated types of anisotropy into account,

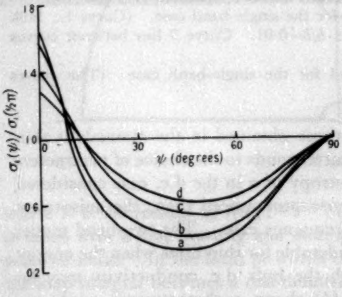


Figure 1.

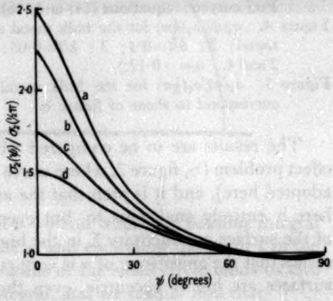


Figure 2.

Figure 1. $\sigma_1(\psi)/\sigma_1(\frac{1}{2}\pi)$ for a thin film as a function of ψ for different values of the ratio thickness/free path. ($a=-0{\cdot}12$, $a'=5$, $a''=-0{\cdot}95$, $l'/l''=0{\cdot}41$; $b/l''=0{\cdot}01$ (curve a), 0·05 (b), 0·2 (c), 0·4 (d).)

Figure 2. $\sigma_2(\psi)/\sigma_2(\frac{1}{2}\pi)$ for a thin film as a function of ψ for different values of the ratio thickness/free path. (The curves correspond to those of figure 1.)

such elaborate refinements seem hardly justified at the present stage of development). In figure 3 we compare, for $b/l'' = 0{\cdot}01$, the exact results obtained from (14) and (16) with the predictions of the approximate formulae (17) and (18) which hold for very thin films, and it is seen that, for such thin films, the simple approximate formulae give reasonably accurate results. In figures 4 and 5, finally, we show for comparison the corresponding results obtained for the bulk metal (using equation (3)) and for thin films when only a single band is considered (using equation (21) and the corresponding expression for σ_2 obtained from (16)).

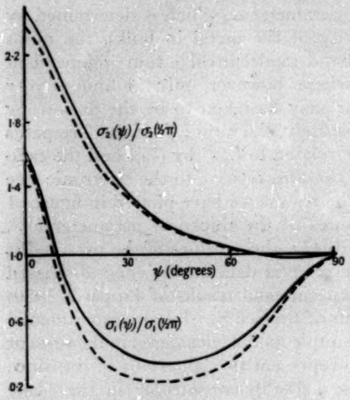


Figure 3.

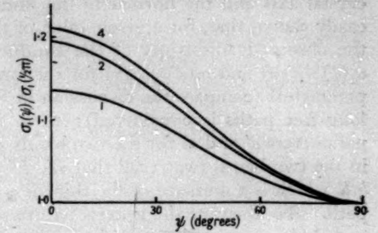


Figure 4

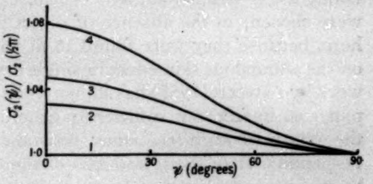


Figure 5

Figure 3. Comparison of exact and approximate formulae for very thin films ($b/l'' = 0{\cdot}01$). Full curves: equations (14) and (16); broken curves: equations (17) and (18).

Figure 4. $\sigma_1(\psi)/\sigma_1(\frac{1}{2}\pi)$ for the bulk metal and for the single-band case. (Curve 1: bulk metal; 2: $b/l = 0{\cdot}4$; 3: $b/l = 0{\cdot}05$; 4: $b/l = 0{\cdot}01$. Curve 3 lies between curves 2 and 4. $a = -0{\cdot}12$.)

Figure 5. $\sigma_2(\pi)/\sigma_2(\frac{1}{2}\pi)$ for the bulk metal and for the single-band case. (The curves correspond to those of figure 4.)

The results are to be compared with those obtained in the anomalous skin effect problem (S, figure 2, where curve 4 corresponds to the choice of parameters adopted here), and it is seen that the anisotropy of σ in the d.c. case considered here is entirely analogous to, but even more pronounced than, the anisotropy of the surface conductivity Σ in the high-frequency case. The two-band model shows that the anisotropy of σ is very considerable for thin films when the energy surfaces are highly eccentric, even though the bulk d.c. conductivity may be almost isotropic, and the conductivity $\sigma_1(\psi)$ shows a characteristic minimum when plotted against ψ which corresponds to the minimum found in $\Sigma_1(\psi)$ in the anomalous skin effect problem. Further discussion of the detailed predictions of the model studied here, and of its adequacy or otherwise as a basis for representing the experimental facts in real metals, must be postponed until such experimental facts have become available.

APPENDIX

We wish to establish the relation (22), i.e. to prove that

$$G(y)\equiv\frac{3}{\pi}\int_1^\infty I\left(\frac{2s}{y}\right)\frac{\sin^2 s}{s^2}\,ds=1-\frac{3}{4y}+\frac{3}{y}\int_1^\infty\frac{e^{-ty}}{t^5}\,dt \qquad (y>0) \qquad \ldots\ldots\text{(A 1)}$$

where $I(u)=2u^{-3}(u-\tan^{-1}u)$, that branch of $\tan^{-1}u$ being chosen which vanishes at $u=0$. Writing $\sin^2 s=\frac{1}{2}(1-\cos 2s)$ and noting that $I(u)$ is an even function of u, we have

$$\begin{aligned}G(y)&=\frac{3}{4\pi}\int_{-\infty}^{\infty}I\left(\frac{2s}{y}\right)\frac{1-e^{2is}}{s^2}\,ds\\&=\frac{3}{2\pi y}\int_{-\infty}^{\infty}I(u)\frac{1-e^{iyu}}{u^2}\,du,\end{aligned} \qquad \ldots\ldots\text{(A 2)}$$

where the principal value of the integrals is understood. The integrand has a simple pole at $u=0$, and is otherwise regular and single-valued in the complex u-plane cut from $-i\infty$ to $-i$ and from $+i$ to $+i\infty$; also $I(u)$ is $O(|u|^{-2})$ for large $|u|$. Hence, if we consider the integral (A2) taken round the closed contour shown in figure 6 we have, in the limit in which the radius of the large semicircle tends to infinity and the radius of the small semicircle tends to zero,

$$G(y)=\frac{3}{2\pi y}\,i\pi R+\frac{3}{2\pi y}\int_{i\infty}^{(i+)}I(u)\frac{1-e^{iyu}}{u^2}\,du \qquad \ldots\ldots\text{(A 3)}$$

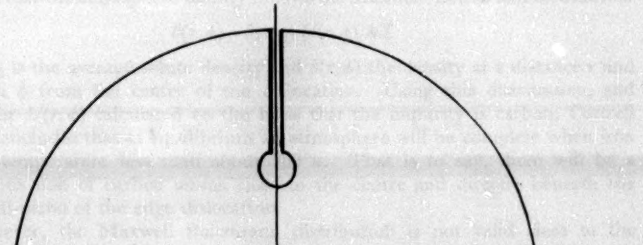


Figure 6.

where R is the residue of $I(u)(1-e^{iyu})/u^2$ at $u=0$ and where the contour integral is taken over a loop starting and ending at $i\infty$ and encircling the point i in the positive sense. In the limit in which the radius of the small circle tends to zero the loop integral becomes a real definite integral, and we readily find that

$$\begin{aligned}G(y)&=1+\frac{3i}{2\pi y}\int_1^\infty\frac{2\pi i}{t^3}\,\frac{1-e^{-yt}}{t^2}\,dt\\&=1-\frac{3}{4y}+\frac{3}{y}\int_1^\infty\frac{e^{-ty}}{t^5}\,dt\end{aligned}$$

which is the required result.

REFERENCES

ANDREW, E. R., 1949, *Proc. Phys. Soc.* A, **62,** 77.
CHAMBERS, R. G., 1952, *Proc. Roy. Soc.* A, **215,** 481.
FAWCETT, E., 1955, *Proc. Roy. Soc.* A, **232,** 519.
HAM, F. S., and MATTIS, D., 1955, *Tech. Rep.* No. 4, Dept of Physics, University of Illinois.
KOURGANOFF, V., 1952, *Basic Methods in Transfer Problems* (Oxford : Clarendon Press).
MACDONALD, D. K. C., and SARGINSON, K., 1950, *Proc. Roy. Soc.* A, **203,** 223.
PIPPARD, A. B., 1950, *Proc. Roy. Soc.* A, **203,** 98 ; 1954, *Ibid.*, **224,** 273.
SONDHEIMER, E. H., 1954, *Proc. Roy. Soc.* A, **224,** 260.

The Effect of Temperature on Cottrell Atmospheres

By N. LOUAT

Aeronautics Research Laboratory, Melbourne, Australia

MS. received 24th May 1955, *and in amended form 22nd November* 1955

Abstract. The variation with temperature of solute atom density close to (within ~10 Å) an edge dislocation is examined. The results of this analysis are used to predict the effect of temperature on the yield point. Good agreement is found between theoretical prediction and three apparently dissimilar experimental results.

---

§ 1. INTRODUCTION

COTTRELL (1948) and his co-workers (1949) have shown that atmospheres of impurity atoms will tend to form around edge-type dislocations. On this basis they have given an explanation for yielding and strain ageing phenomena in iron and other metals. In formulating these theories it has been assumed that the atmosphere density follows the Maxwell–Boltzmann distribution

$$\theta(r,\phi)=\theta_0 \exp U(r,\phi)/\boldsymbol{k}T$$

where θ_0 is the average solute density and $\theta(r,\phi)$ the density at a distance r and direction ϕ from the centre of the dislocation. Using this distribution, and values for $U(r,\phi)$ calculated on the basis that the impurity is carbon, Cottrell (1953) concludes that at equilibrium an atmosphere will be complete when iron is at a temperature less than about 700°K. That is to say, there will be a continuous line of carbon atoms close to the centre and directly beneath the extra half-plane of the edge dislocation.

However, the Maxwell–Boltzmann distribution is not valid close to the dislocation since Cottrell and Bilby (1949) have shown that only one solute atom per atom plane of dislocation should lie within about 10 Å of the centre. The purpose of the present paper is to take some account of this fact and to examine its consequences.

§ 2. ATMOSPHERE DENSITIES

The treatment adopted here is essentially qualitative and a simplified model is used. It is supposed, considering one atom plane of dislocation, that the region near a dislocation may be divided into a number of discrete sub-regions of various sizes, each characterized by a unique solute atom–dislocation interaction energy, and that at most only one solute atom can occupy each sub-region at any instant. It is assumed further that the interaction between solute atoms in neighbouring sites along the length of the dislocation may be neglected. For an assembly of atom planes under these conditions it may be shown, either by setting up the appropriate partition functions, or more simply by considering directly the thermally activated equilibrium interchange of atoms between adjoining energy levels, that the resultant distribution is

$$\frac{n_1}{n_0}=\frac{N_1-n_1}{N_0-n_0}\exp\left(\frac{U_1-U_0}{\boldsymbol{k}T}\right)$$

or

$$\frac{\theta_1}{\theta_0}=\frac{1-\theta_1}{1-\theta_0}\exp\left(\frac{U_1-U_0}{\boldsymbol{k}T}\right)$$

where N_1 and N_0 are the number of possible positions for a solute atom of free energy U_1 and U_0 respectively in a dislocation which is N atomic planes in length. $n_1=N_1\theta_1$ and $n_0=N_0\theta_0$ are the number of these positions which are occupied. It may be seen that this distribution reduces to the Maxwell–Boltzmann form when θ_1 is small. (Lomer has apparently reached similar conclusions (see Cottrell 1953 b).)

Further, for small θ_0 and with $U_0=0$, the density in the lowest energy level (energy V) is given by

$$\theta=\frac{\theta_0\exp(V/\boldsymbol{k}T)}{1+\theta_0\exp(V/\boldsymbol{k}T)}$$

and θ tends to an equilibrium value of 1 when $\theta_0\exp(V/\boldsymbol{k}T)\to\infty$ and hence as $T\to 0$ provided $\Sigma N_i\theta_i\geqslant N$.

However, the time taken for a state of equilibrium to be reached depends on the rate of diffusion of solute atoms in the parent lattice. Since this rate is usually low at room temperature and below, it follows that atmosphere densities in material used in tests at low temperatures will have been determined by the temperature at which significant diffusion last took place. This fact can be expected to be of practical importance only when $\theta_0\exp(V/\boldsymbol{k}T)\sim 1$ or less, where T is some elevated temperature, and the rate of cooling to the testing temperature is sufficiently rapid to prevent considerable diffusion during cooling. It is doubtful whether these conditions can be satisfied without special precautions. Finally, it should be noted that, under equilibrium conditions, θ_0 is determined by the solid solubility of the solvent and is, in general, a monotonic function of the temperature.

§ 3. THE EFFECT OF TEMPERATURE ON A PREDETERMINED COTTRELL ATMOSPHERE

We shall now investigate how such a predetermined atmosphere would behave with varying temperature, and the consequent effect on the mechanical properties of the metal.

Cottrell (1948) gives, with a change in sign, the interaction energy of a solute atom with an edge dislocation as

$$U(r,\phi)=-\Delta v\frac{\mu b}{3\pi}\,\frac{1+\nu}{1-\nu}\,\frac{\sin\phi}{r}=\frac{-A}{r}\sin\phi \qquad \ldots\ldots(1)$$

where Δv is the dilatation produced by the solute atom, μ the shear modulus and ν Poisson's ratio for the lattice; b is the magnitude of the Burgers vector and ϕ the angle between the direction of the Burgers vector and a line joining the solute atom to the centre of the dislocation.

Cottrell and Bilby (1949) estimate for the case of carbon in iron that $A\simeq 3\times 10^{-20}$ dyn cm^2, but consider that such a value leads to a gross over-estimate of the interaction energy when the solute atom is only a few ångström units from the dislocation centre.

Now, since the expected error in A will be greater when r is small, we must anticipate that the curve $[U(r,\phi),r]$ will be appreciably flatter for small r than would be predicted from equation (1). To take account of this, we shall suppose that there exists a small region contiguous with the dislocation centre in which the interaction energy is sensibly constant. We suggest further that in this region, which in the case of interstitial solute atoms is one of extreme dilatation, the activation energy for movement of solute atoms from one site to another is much less than in the bulk of the crystal.

It is to be expected, then, that the equilibrium distribution of solute atoms within such a region would vary with temperature and that the time taken to reach equilibrium would be small, even at low temperatures. For simplicity we assume that each region (again one atomic distance in length) may be divided into two parts, each portion being characterized by a unique value of the interaction energy. Further, we consider that the portion remote from the dislocation centre contains β times as many solute sites as the central volume. Now, for temperatures at which diffusion is not important in the bulk of the crystal, we may, in accord with our earlier conclusions, consider that the total number of effective solute atoms in a length of dislocation is constant and equal to $N\theta_0 \leqslant N$ where N is the total number of atom planes in the dislocation under consideration. It may easily be shown then, that the density in the central portion is

$$\alpha=\frac{\theta_0 \exp(W/\boldsymbol{k}T)}{\beta+\exp(W/\boldsymbol{k}T)} \qquad \ldots\ldots(2)$$

where W is the difference in energy of the two levels.

§ 4. Effect on Mechanical Properties

4.1. *Dislocation Locking*

In accordance with the concept of an 'equipotential' region we cannot expect that equation (1) will be valid for small r. Therefore we cannot, without further examination, consider, as in effect do Cottrell and Bilby, that the restoring force on a dislocation displaced laterally a distance x, parallel to the Burgers vector, from the solute atom is given by

$$\frac{\partial}{\partial x}\frac{A\sin\phi}{r}=\frac{A\sin 2\phi}{r^2}$$

where r is small.

Further, we note from the work of Peierls (1940), Foreman, Jaswon and Wood (1951), and others, that a dislocation in a crystal lattice must be considered to have considerable lateral extent which in the case of the edge dislocation is parallel to its Burgers vector.

As a first step in determining the influence of solute atoms on mechanical properties we have to enquire what effect the presence of solute atoms will have on the dislocation distribution and, since we are mainly interested in condensed atmospheres, we have to consider the results of the presence of numbers of solute atoms arranged in what is approximately a straight line parallel to the dislocation. The shear stresses from such a line of solute atoms centred at a point $-\delta, y$ and lying parallel to the z axis are

$$\sigma_{xy}\simeq\frac{2\mu\Delta v}{\pi a}\,\frac{(x+\delta)y}{\{(x+\delta)^2+y^2\}^2}$$

where Δv is the volume change in the crystal caused by the introduction of a solute atom and a is the atomic spacing along the line of the solute atoms. Following Peierls one may write an equation for the dislocation distribution when such shear stresses are present. We have

$$\frac{\mu}{2\pi(1-\nu)}\int_{-\infty}^{\infty}\frac{du}{dx}\,\frac{d\bar{x}}{\bar{x}-x}=\frac{-b\mu}{2\pi d}\sin\left(\frac{4\pi u}{b}\right)-\frac{2\mu\Delta v}{\pi a}\,\frac{(x+\delta)y}{\{(x+\delta)^2+y^2\}^2}$$

where u is the displacement of atoms on the surface of the upper half crystal from their normal positions. No solution has as yet been obtained for this equation but it seems reasonable to assume that the effect of the solute atoms, when close to the dislocation centre, is to make the Peierls approximation very nearly correct since the maximum values of the two parts of the right-hand side of the above equation are of the same order of magnitude, supposing, as Cottrell and Bilby do, that $y=2\,\text{Å}$ and that (for carbon in iron) $\Delta v=7{\cdot}8\times10^{-24}\,\text{cm}^3$. Adopting this view and restricting ourselves to those cases where x is small, we shall suppose that the dislocation width $\zeta(y)$, for small y, is some monotonic function of y only and consider (Leibfried and Lucke 1949) that the stresses in the lower half crystal behave as if the dislocation were a line singularity located at $x=0$, $y=\zeta$. We find that equation (1) becomes

$$u(\bar{r},\bar{\phi})=\frac{-A[y+\zeta(y)]}{x^2+[y+\zeta(y)]^2}$$

where $\bar{r}$, $\bar{\phi}$ refer to the new centre of coordinates $(0,\zeta)$. Also, it may be seen that the elastic energy of the dislocation will increase as ζ decreases and it is found by evaluating the energy contained in two infinite regions bounded by planes normal to the y axis which pass through the points $y=\zeta_1$, and $y=\zeta_2$ in one case and $y=-\zeta_1$, $y=-\zeta_2$ in the other, that the increase in elastic energy per atomic length of dislocation due to a change in ζ from ζ_1 to ζ_2 is given by

$$\mu_\zeta\simeq\frac{\mu b^3}{4\pi(1-\nu)}\ln\left(\frac{\zeta_1}{\zeta_2}\right).$$

Substituting values suitable for iron one finds that $\mu_\zeta\sim0{\cdot}5$ ev per atom plane when $\zeta_1\sim1{\cdot}6\zeta_2$, which seems a not unreasonable change in ζ to be caused by the movement of the solute atom (see Foreman, Jaswon and Wood 1951).

A tendency for the existence of an 'equipotential' region may then be inferred since the solute atom–dislocation interaction energy is the sum of two quantities $U(r,\phi)+\epsilon$ and U_ζ whose maxima are of similar magnitude but of opposite sign. Here ϵ represents the energy of other, apparently minor, interactions including electrical interaction (Cottrell, Hunter and Nabarro 1953), and those due to the presence of shear stresses (Crussard 1950).

We shall assume that such an 'equipotential' region does exist and also consider that the restoring force per atom plane is given by

$$\frac{\partial}{\partial x}\left[\frac{A\sin}{\bar{r}}\bar{\phi}+U_\zeta\right]\simeq2Ax(\zeta+y)/\{(\zeta+y)^2+x^2\}$$

with the assumption that ζ does not change significantly with x for small values of x, which seems reasonable.

A result of the same form but differing by a numerical factor is obtained if we consider the interaction between the shear stresses due to a row of solute atoms and a dislocation distributed according to the Peierls relation, assuming as before that ζ remains constant for the small displacements involved. In the former case the maximum restoring stress with $y=0$, $\zeta=2$ Å and $A=3\times10^{-20}$ dyn cm^{2} (see Cottrell and Bilby 1949) is 2×10^{11} dyn cm^{-2}. The second method gives $\sigma_0\simeq6\times10^{10}$ dyn cm^{-2} with the same values for y and ζ. Of these, the latter result is perhaps to be preferred since in this case we have considered the effect of a row of solute atoms which is a better representation of condensed atmosphere than is a single solute atom.

Again, we see that the restraining stresses for the largest values of y, perhaps 2–3 Å, in the 'equipotential' region are likely to be markedly different from those for the smallest values of y since both y and ζ should increase together in this range. Thus, in accordance with the original approximation of dividing the 'equipotential' region into two discrete interaction energy levels, we may characterize these levels by two widely different values of the maximum restraining stress, e.g. σ_0 for $y=0$, $\zeta=2$ Å, and $\sigma_0/16$ for $y=3$ Å, $\zeta=5$ Å.

Finally, it must be remarked that while little reliance can be placed on the quantitative results of this analysis, the qualitative information should indicate the trends in behaviour to be expected in real crystals and thus provide sufficient grounds to justify investigating the consequences of Cottrell atmospheres having the structure we have envisaged. Thus, in the next section we shall use this model to predict the variation of the yield stress with temperature.

4.2. *Temperature Dependence of the Yield Stress*

The behaviour under applied shear stresses of great lengths of dislocation may now be determined by finding the maximum expected length which contains solute atoms in an outer, but not in an inner position.

We shall neglect the effect of thermal stresses on 'bound' lengths of dislocation for several reasons. Thus, Cottrell and Bilby in their analysis of this question consider that the thermally supplied activation energy U varies in such a fashion that $U/\boldsymbol{k}T$ is a constant, and that this constant should not exceed about 50. To achieve this value they found it necessary to set the energy per atomic length of dislocation as 1 ev and their factor A as 3×10^{-21} dyn cm^{-2}. But Cottrell's (1953 a) analysis of Dijkstra's experiments indicates that $V\simeq0{\cdot}5$ ev, which, using the Cottrell–Bilby relation, gives $A=1{\cdot}5\times10^{-20}$ dyn cm^{-2}, which is in rather better agreement with their theoretically derived value ($A=3\times10^{-20}$ dyn cm^{-2}). Again, it is felt that 1 ev per atom plane of dislocation is too small an estimate when it is remembered that μb^3 for iron is 8·75 ev. Thus, if $A=1{\cdot}5\times10^{-20}$ dyn cm^{-2} and the energy per atom plane of dislocation is x ev, Cottrell and Bilby's equation for U leads to $U/\boldsymbol{k}T=105\sqrt{x}$.

Further, Cottrell and Bilby in evaluating U take no account of the elastic strain energy associated with the thermal stress in the surrounding crystal which leads to the bending of the dislocation. This energy would appear to be appreciable (Becker 1925) if σ_0, the stress equivalent of the maximum force exerted by the solute atoms, is as large as 6×10^{10} dyn cm^{-2}, which is the most recent estimate (Cottrell 1953b).

These considerations are not sufficient to invalidate the Cottrell–Bilby mechanism, but it is considered that they raise sufficient doubts to justify separate consideration of another process.

Following Koehler (1952), and assuming random distribution, we have the probability of finding two locking points separated by l/b atoms in a great length of N atoms is $p=\alpha(1-\alpha)^{l/b}$ where $\alpha=k/N$ is the concentration of locking points and k is the number of them. Hence, the number of lengths between l and $l+\delta l$ is

$$N(l)\,\delta l=N(1-\alpha)^{l/b}\,[\ln(1-\alpha)]^2(1-\alpha)\frac{\delta l}{b}$$

and the number greater than or equal in length to a certain length L is

$$-N(1-\alpha)\ln(1-\alpha)(1-\alpha)^{L/b}.$$

Now, it is suggested that, for a sharp yield point to be observed, there need be only one such length greater than or equal to a certain critical length in a region of the crystal in which the maximum shear stresses occur. It is supposed that this region contains MN atomic lengths of dislocation. Hence it is required that

$$-MN(1-\alpha)\ln(1-\alpha)(1-\alpha)^{L/b}=1$$

whence

$$\frac{b}{L}\simeq\frac{\ln[1/(1-\alpha)]}{\ln[MN(1-\alpha)]}. \qquad \ldots\ldots(3)$$

Now, under the action of a sufficiently large stress, a length L will tear away from the restraining forces (principally) at its ends. Thereafter, the increased length of free dislocation will continue to grow longer until the entire length of dislocation is free and able to operate as a Frank–Read source. The condition for initial movement is approximately

$$\theta Lb\frac{\sigma_0}{n}+2\sigma_0 b^2=L\sigma b \qquad \ldots\ldots(4)$$

where σ_0 is the shear stress equivalent to the maximum restraining force, σ the applied shear stress and n a dimensionless constant whose value is the ratio between the maximum restraining forces of the lower and upper levels of the 'equipotential' region. The first term of this equation represents the effect of those atoms which lie in the upper level and which are considered to exert a uniform drag $\theta_0\sigma_0/n$ per unit length over the whole 'free' length, where θ_0 is the solute density over the 'free' length.

Equation (4) reduces to

$$\frac{\sigma}{\sigma_0}=\frac{2b}{L}+\frac{\theta_0}{n}$$

and combining this result with equations (2) and (3) we have

$$\frac{\sigma}{\sigma_0}=\left[2\ln\left\{\frac{\beta+e^{W/kT}}{\beta+(1-\theta_0)e^{W/kT}}\right\}\Big/\ln[MN(1-\alpha)]\right]+\frac{\theta_0}{n}+\delta$$

where δ is a term, independent of temperature, added to allow for the internal stresses etc. which Cottrell (1950) has shown can cause an increase in the yield point.

§ 5. COMPARISON WITH EXPERIMENT

In view of the number of disposable constants available, it is not surprising that this function can be made to fit, with sufficient accuracy, the experimental observations of McAdam and Mebs (1943). These observations are compared, in figure 1, with the equation

$$\sigma=1\cdot5\times10^9\ln\left\{\frac{3+e^{590/T}}{3+e^{-4.6}e^{590/T}}\right\}\text{ dyn cm}^{-2}.$$

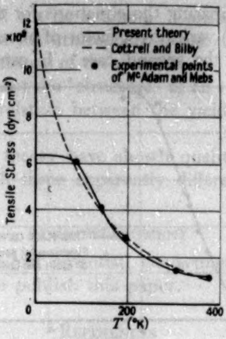


Figure 1.

Here, in order to obtain a solution in a reasonable time, the constant terms and the term in n, which appear to be small, were ignored. In view of this and other approximations made earlier, no particular significance can be attributed to the values obtained for the parameters. However, for the sake of clarity typical values are given in the following table.

T	L	α	β	θ
303	$84b$	0·7	3	0·99
144	$29b$	0·95	3	0·99

It may be seen that the shape of the predicted curve at temperatures below those employed by McAdam and Mebs differs from that suggested by Cottrell and Bilby (1949). This low temperature prediction is supported to some extent by other work. Thus, our general expression may be fitted, employing the values for β and W found from the data of McAdam and Mebs, to the results obtained by Eldin and Collins (1951) from hot rolled SAE 1020 steel.

The calculated curve and experimental points are shown in figure 2. The equation employed is

$$\sigma = 0{\cdot}89 \times 10^9 \ln\left\{\frac{3 + \mathrm{e}^{590/T}}{3 + \mathrm{e}^{-10{\cdot}1}\mathrm{e}^{590/T}}\right\} + 3{\cdot}4 \times 10^9 \text{ dyn cm}^{-2}.$$

It may be seen that the factor $1 - \theta_0$ is considerably smaller here than is the case for the previous result, and it is apparent that, over much of the range of temperature represented, $(1-\theta_0)\exp(W/\boldsymbol{k}T) \ll \beta$ while at the same time $\exp W/\boldsymbol{k}T \gg \beta$. Under these conditions we should expect that σ/σ_0 would be proportional to $1/T$. It was to illustrate that such a limited linear dependence does occur that the yield stresses were plotted against $1/T$. In addition it may be seen that flattening predicted at low temperatures does in fact occur. However, it must be remembered that Eldin and Collins did not observe a yield point at temperatures below 61°K and the points plotted in this range represent the onset of fracture. This is not a serious objection if one considers, as suggested by Zener (1948), that fracture occurs only after some plastic deformation involving substantial movement of dislocations.

It may be seen by comparing the equations for the two sets of experiments that the term $\ln MN(1-\alpha)$ varies by a factor of about 1·5 and that the constant term is not negligible in the second case. It cannot be claimed that these differences are understood, but it is felt that they are not unreasonable in view of the different nature and history of the materials employed in the two cases.

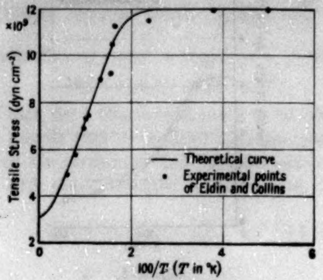


Figure 2.

Finally, one notes that if $\beta \gg \exp(W/\boldsymbol{k}T)$ the general expression reduces to

$$\frac{\sigma}{\sigma_0} = \frac{2\theta_0 \exp(W/\boldsymbol{k}T)}{\beta \ln[MN(1-\alpha)]} + \frac{\theta}{n} + \text{constant}.$$

This expression is to be compared with the results of Rosi and Mathewson (1950) for 99·996% aluminium tested in tension. These workers found that $\sigma \simeq 6{\cdot}1 \times 10^6 \exp(150/T)$ dyn cm^{-2} where σ is the stress at which plastic deformation first became appreciable. A similar temperature dependence has been found by Schmid (1931) for the same property in zinc.

It is apparent that, unless β is much greater than the value ($\sim$3) found for the case of steel, no such exponential dependence on $1/T$ can be expected. But it is possible that β may satisfy the required condition. Thus, in a face-centred cubic or hexagonal-close-packed metal, Cottrell locking will, in the absence of interaction between solute atoms and shear stress fields of the dislocation, be least for screw dislocations and will require the dissociation of complete dislocations as described by Heidenreich and Shockley (1948). In the case of an originally pure screw dislocation, each partial dislocation has a Burgers vector whose edge component has a magnitude b.

Now Cottrell's quantity A is proportional to the product of the Burgers vector and the shear modulus so that, for the case of a screw dislocation in aluminium, A and hence V have only about one-tenth of their values in iron for the same solute dilatation of the lattice. It is likely, therefore, that the magnitude of β will be determined by the size of the region within which diffusion processes can occur easily. β may, on this account, be considerably greater than is the case for carbon atoms in iron.

§ 6. CONCLUSIONS

It has been found that the atmosphere density is characterized by a particular equilibrium value determined by the density of solute and the ambient temperature. In general, when a metal containing solute atoms which interact with dislocations has been cooled to a temperature at which the solute diffuses slowly, this equilibrium will not be attained. Consequently, those mechanical properties which depend on the atmosphere density will also depend to some extent on the temperature history of the material.

Again, it has been seen that the 'structure' of an atmosphere is temperature-dependent and leads to a relation between the yield strength and the testing temperature.

The predictions of the theory have shown qualitative and to some extent quantitative agreement with three apparently different types of experimental results.

ACKNOWLEDGMENT

Thanks are due to the Chief Scientist, Department of Supply, Melbourne, Australia, for permission to publish this paper.

REFERENCES

BECKER, R., 1925, *Phys. Z.*, **26,** 919.
COTTRELL, A. H., 1948, *Report of a Conference on Strength of Solids* (London : Physical Society) p. 46 ; 1950, *Symposium on the Plastic Deformation of Solids*, Pittsburgh, p. 60 ; 1953 a, *Dislocations and Plastic Flow in Crystals*, Oxford ; 1953 b, *Progr. Metal Phys*, **4,** 240.
COTTRELL, A. H., and BILBY, B. A., 1949, *Proc. Phys . Soc.* A, **62,** 49.
COTTRELL, A. H., HUNTER, S. C., and NABARRO, F. R. N., 1953, *Phil. Mag.*, **44,** 1064.
CRUSSARD, C., 1950, *Métaux Corros-Usure*, **25,** 203.
ELDIN, A. S., and COLLINS, S. C., 1951, *J. Appl. Phys.*, **22,** 1296.
FOREMAN, A. J., JASWON, M. A., and WOOD, J. K., 1951, *Proc. Phys. Soc.* A, **64,** 156.
HEIDENREICH, R. D., and SHOCKLEY, W., 1948, *Report of a Conference on Strength of Solids* (London : Physical Society) p. 57.
KOEHLER, J. S., 1952, *Phys. Rev.*, **86,** 52.
LEIBFRIED, G., and LUCKE, K., 1949, *Z, Phys.*, **126,** 450.
MCADAM, D. J., and MEBS, R. W., 1943, *Proc. Amer. Soc. Test. Mater.*, **43,** 661.
NABARRO, F. R. N., 1952, *Advanc. Phys., Phil. Mag.* Suppl., **1,** 271.
PEIERLS, R. E., 1940, *Proc. Phys. Soc.*, **52,** 23.
ROSI, F. D., and MATHEWSON, C. H., 1950, *J. Metals, N.Y.*, **188,** 1159.
SCHMID, E., 1931, *Z. Elecktrochem.*, **37,** 447.
ZENER, C., 1948, *Fracturing of Metals* (Cleveland, Ohio : American Society for Metals) p. 3.

The Positive Column in a Longitudinal Magnetic Field

BY R. J. BICKERTON AND A. VON ENGEL
Clarendon Laboratory, University of Oxford

MS. received 30th June 1955, and in amended form 25th November 1955

Abstract. A magnetic field parallel to the axis of the positive column of a discharge is known to impart a rotary component on the radial motion of electrons. This reduces the flow of both positive ions and electrons to the wall, causing a corresponding decrease in the radial and longitudinal components of the electric field and in the electron temperature.

It is shown here that under certain conditions Langmuir probes can be applied to explore a plasma in a magnetic field. A new method of measuring the concentration of charges is used: first relative values of ion concentration are obtained from the ion saturation currents in various magnetic fields. This result together with the electron saturation currents in zero magnetic field gives absolute values of concentration of charges in a magnetic field. Such measurements have been carried out in helium at low pressures.

The effect of the magnetic field on the radial distribution of charges, the rate of ionization, the electric field, the electron temperature, the average energy loss in a collision, and the radial potential distribution is discussed. It is concluded that in zero magnetic field the Langmuir theory of free ion fall describes best the properties of the discharge plasma whereas in a magnetic field of sufficient strength Schottky's theory of ambipolar diffusion applies.

When the gas becomes highly ionized, the effect of the partial pressure of the electron gas may become important so that finally both the longitudinal component of electric field and the electron temperature should become independent of the magnetic field. Previous observations of a low pressure positive column in caesium at high current densities seem to support this view.

---

§ 1. INTRODUCTION

When a steady uniform positive column of a low pressure discharge is acted upon by a longitudinal magnetic field, the charged particles, having velocity components in all directions, spiral about the magnetic lines of force. Because of their small mass, only the electrons are appreciably affected by the magnetic field. The spiralling parallel to the axis of the tube between collisions reduces the radial diffusion of the electrons and thus a smaller radial electric field is required to maintain the equality between the numbers of ions and electrons arriving at the non-conducting tube wall. Since their radial velocities are the same, the radial flow of both charges will decrease. A longitudinal magnetic field should thus reduce the electron temperature and the electric field in the column.

In an attempt (Tonks 1939) to formulate a quantitative theory, use was made of a theoretical result (Townsend 1915) for the effect of a magnetic field on electron diffusion in the absence of space charge. This result was confirmed in experiments with electron swarms for photoelectric currents (Bailey 1930). It holds also for larger currents ($\leqslant 10^{-6}$ amp) when allowance is made for the space charge. Other experiments (Cummings and Tonks 1941) with a positive column in mercury vapour were inconclusive because of the difficulty of interpreting probe characteristics taken in the presence of a magnetic field. Recently Davies (1953) used a spectroscopic method to investigate the influence of a magnetic field on the electron temperature in a low pressure positive column in caesium vapour. Instead of a large reduction, however, a small increase in the electron temperature was observed. The effect of a magnetic field on the loss of charges from a plasma to the walls of a closed conducting box has been shown by Simon (1955) to be controlled by currents in the plasma which are flowing essentially along the field and not by plasma oscillations as previously suggested. These arguments do not apply here because the plasma of a positive column is uniform in the direction of the magnetic field.

In this paper probe measurements in a positive column in helium in longitudinal magnetic fields are presented and the effect of such fields on the column is discussed theoretically.

§ 2. THEORY OF PROBES IN A MAGNETIC FIELD

The theory of probes in zero magnetic field (Langmuir and Mott-Smith 1924) rests on two assumptions, viz.

(*a*) The dimensions of the probe are small compared with the mean free path of ions and electrons. This means that the probe can be assumed to collect only a relatively small number of charged particles in the plasma around the sheath, so that outside this sheath the plasma is, to a close approximation, undisturbed by the presence of the probe.

(*b*) The thickness of the space charge sheath surrounding the probe is small compared with the mean free path of ions and electrons. Thus the sheath can be treated as a region in which ions and electrons move as in vacuum, undisturbed by collisions. Its thickness $d \propto V_{\mathrm{p}}^{3/4}/n^{1/2}$ for constant mean ion energy, where V_{p} is the potential across the sheath and n the ion concentration.

A magnetic induction B applied to the plasma effectively reduces the free paths of the charges in directions perpendicular to B to less than the radius of path curvature $\rho = v/\omega = mvc/eB$, v being the velocity and m the mass of the particle. With a probe collecting across the magnetic field assumption (*a*) will be invalid in moderate magnetic fields, e.g. $\rho_{\mathrm{e}} = 4 \times 10^{-2}$ cm for 1 ev electrons and $B = 100$ gauss. The validity of (*b*) depends on the sheath thickness d and thus on the plasma density n, the type of gas and on B. If (*b*) holds, the motion of the charged particles is complicated by the presence of the magnetic field. When (*b*) is not valid, electrons make collisions in the sheath.

In our experiments in He, the ions are relatively light, the plasma density low ($n \simeq 10^{10}$ cm^{-3}) and thus the sheath thickness for negative probe potentials is large ($d \simeq 1$ mm). The electrons therefore collide many times with gas molecules in crossing the sheath. The number of electrons striking the negative probe per second is a small fraction of the electrons entering the sheath from the plasma; the majority return to the plasma. Thus to a first approximation the net flow across the sheath can be neglected and the spatial distribution of the electrons follows Boltzmann's law for particles diffusing in a potential field, assuming that the velocity distribution of the electrons is Maxwellian. The presence of the magnetic field does not invalidate this law (Chapman and Cowling 1939). Thus,

$$n_\mathrm{p} = n_0 \exp(-eV_\mathrm{p}/kT_\mathrm{e}) \qquad \ldots\ldots(1)$$

where n_p and n_0 are the number density of electrons at the surface of the probe and the plasma–sheath boundary respectively, and T_e the electron temperature.

It can be shown (Bickerton 1954) that the number of electrons striking 1 cm² per second of a solid plane parallel to B decreases as B rises, viz.

$$\frac{j}{e} = \frac{n\bar{c}}{4}\left[\frac{8+(\omega\tau)^2\{1-\exp(-2\pi/\omega\tau)\}}{2\{4+(\omega\tau)^2\}}\right] = \frac{n\bar{c}}{4} f(B). \qquad \ldots\ldots(2)$$

The mean collision time $\tau = l_\mathrm{e}/\bar{c}$, where $\bar{c}$ is the mean random velocity and l_e is the mean free path of the electrons. The number of electrons crossing in one direction per second 1 cm² of an imaginary plane in the plasma is $n\bar{c}/4$ independent of B because the electrons move again and again on curved paths through the plane and back. A solid plane however, is a sink for electrons and thus the contribution to the electron flux from those moving back is absent and the number striking the plane is smaller. In an actual discharge n_0, $\bar{c}$ and T_e are functions of B.

From (1) and (2) the current density of electrons to the probe is

$$j = \tfrac{1}{4} e n_0 \bar{c} f(B) \exp(-eV_\mathrm{p}/kT_\mathrm{e}) \qquad \ldots\ldots(3)$$

and so

$$(d \ln j/dV_\mathrm{p})_B = -e/kT_\mathrm{e}. \qquad \ldots\ldots(4)$$

Thus the inverse gradient of the semi-log plot of the electron current is proportional to the electron temperature (as in zero magnetic field) in the region of high negative probe potential.

As the probe potential approaches that of the plasma, the electron current to the probe increases, and the use of the Boltzmann law (1) becomes unjustified. Also the sheath thickness decreases, so that the assumption that the electrons collide frequently in the sheath fails. Further, since the probe is of finite size it draws some of its current along the magnetic induction B and some across B. Thus the term $f(B)$ in (2) which was derived for an infinite plane probe is not strictly applicable. For these reasons it is impossible to determine the number concentration of charges from the saturation electron current in the conventional manner.

In our experiments the number concentration was obtained by measuring the relative changes in the saturation ion current to the probe for various magnetic field strengths, and using these together with the values of the plasma concentration found from the electron saturation current in zero magnetic field. This procedure assumes that the motion of the ions across the sheath is largely unaffected by the magnetic field, which is true here when $B \leqslant 600$ gauss, and that the electron temperature is sensibly constant. The effect of the electron temperature on the ion current even in extreme cases is small (Bohm, see Guthrie and Wakerling 1949).

The theory of probes in a magnetic field in a dense plasma, i.e. when $\rho_\mathrm{e} \geqslant d$, and with heavier ions, is more complex (see also Spiwak and Reichrudel 1938).

§ 3. The Apparatus

The discharge tube used is shown in figure 1. The wall probe consisted of a molybdenum disc, the other probes were all of tungsten, sheathed in glass. The helium gas used was purified before each experiment by maintaining it in contact for several hours with activated charcoal, cooled to the temperature of liquid oxygen.

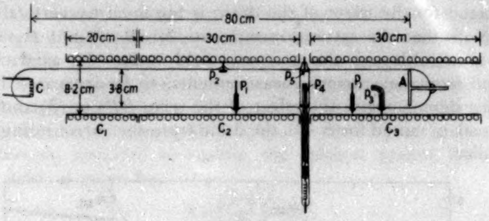


Figure 1. C, hot tungsten cathode; A, copper anode; P_1, floating tungsten probes, used to determine X; P_2, molybdenum disc wall probe (5 mm diameter); P_3, tungsten probe sheathed in Pyrex, collecting along H; P_4, movable tungsten probe, sheathed in Pyrex, collecting across H; P_5, copper ring flush with inside ot tube wall; C_2, C_3, matched pair of coils; C_1, strong field coil to reduce end effects.

Three coils were used to provide the magnetic field (figure 1). Two of these constituted a matched pair, each giving a field of 110 gauss per ampère of coil current. When used with a 1 cm gap between their ends (necessary because of the projecting movable probe), the magnetic field at the centre of this gap was 85% of that at the centre of each coil. The current through the third, and shorter coil, was adjusted so that the magnetic field at its centre was twice that at the centres of the other two coils. In this way end-effects are reduced (Cummings and Tonks 1941).

To determine the radial potential variation in the column a radially movable tungsten wire probe P_4 and a copper ring P_5 on the tube wall were used. The floating voltage V_f between these two probes was measured for various positions of the tungsten probe. From figure 1, and Kirchhoff's laws, then, neglecting contact potentials (see below),

$$V_{sw} + V_p' - V_{sp} - V_f = 0$$

where V_{sw} is the potential across the sheath in front of the ring on the tube wall, V_{sp} is the potential across the sheath in front of the movable probe, and V_p' is the potential drop in the plasma between the two plasma–sheath boundaries. As long as the movable probe remains deep in the plasma, the potentials V_{sp} and V_{sw} may be assumed to be independent of the radial position of the probe. Thus from V_f relative changes in V_p' with position can be determined. Since only relative changes are found, the neglect of contact potentials is justified. When the movable probe is near to the tube wall the two probe sheaths interact, and this makes reliable interpretation of floating potential measurements impossible.

§ 4. Experimental Results

Typical probe curves obtained are shown in figure 2. That taken in a magnetic field is seen to have a linear portion at negative probe potential. The surface of the probe was on the axis of the tube and directed so as to collect current across B. The break in the characteristic for $B=0$ is ill defined so that the plasma potential cannot be found accurately. This is because the probe drains an appreciable number of electrons from the surrounding plasma. V_p has been taken at the point of maximum curvature. An estimate of the saturation electron current corrected for the effect of the drain is obtained by vertical extrapolation as shown. Note that the saturation currents for $B=0$ and $B=440$ gauss are sensibly the same, although the magnetic field causes the axial charge density n_0, determined from ion current measurements, to be increased by a factor of about 4. This demonstrates the effect of the term $f(B)$ in (3) and also explains the sharper break obtained for $B \neq 0$, the drain from the surrounding plasma being relatively smaller.

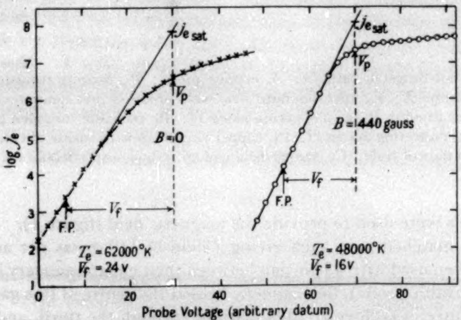


Figure 2. Helium, $p=0{\cdot}086$ mm Hg, V_p=plasma potential, $R=1{\cdot}8$ cm, $i=0{\cdot}1$ A; F.P. = floating potential.

The results obtained from the probe measurements are summarized in figures 3–5. The current i through the tube was 100 mA, its density of order 10^{-2} A cm^{-2}. An example of the effect of the magnetic field on various discharge parameters at a particular gas pressure is given in figures 6–8. The tube radius R was of the order of 2 cm throughout, only the pressure p in the parameter Rp being varied. Electron temperature measurements made with the movable probe showed this temperature to be independent of radial position in the column, both with and without the magnetic field.

The pressure range of the experiments was limited by two factors. Above 1 mm Hg the maximum magnetic field available (600 gauss) has only a negligible effect on the electron temperature. At the lowest pressure used (0·022 mm Hg) the probe characteristics in a magnetic field indicate the presence of two groups of electrons, one having a distinctly higher 'temperature' than the other. This is due to the persistence, in the part of the discharge in the uniform magnetic field, of electrons having a temperature corresponding to the part outside the field, or to the part where the magnetic field is such as to carry electrons to the wall. (The mean free path of the electrons in the direction of B is of the order of 4 cm at this pressure.) The fact that the relative number of electrons in the fast group is found to increase with increasing B indicates that the relaxation distance rises as B rises. Therefore this apparatus is not suitable for measurements at still lower pressures, when a longer discharge tube would be required.

§ 5. DISCUSSION

5.1. *Electron Temperature in Zero Magnetic Field*

The ambipolar theory of Schottky (1924) is applicable to the positive column when the pressure is such that the mean free path of the ions is small compared with the tube radius ($l_+ \ll R$). It is assumed that ionization occurs by single collisions only, and that ions and electrons are lost by recombination on the tube wall. If it is further assumed that the number of electrons at the wall n_w is small compared with that on the axis n_0 then λ, the number of ion pairs produced per second per electron required to sustain the column against loss by diffusion in the electric field, is given by

$$\lambda = \left(\frac{2{\cdot}4}{R}\right)^2 D_a \qquad \ldots\ldots(5)$$

where D_a is the ambipolar diffusion coefficient defined by the equation

$$D_a = \frac{D_e D_+ (T_+ + T_e)}{D_e T_+ + D_+ T_e} \simeq D_+ \frac{T_e}{T_+} \qquad \ldots\ldots(6)$$

where + refers to positive ions.

The linearity of the lower part of the probe curves (figure 2) suggests a Maxwell distribution of electron speeds. For a Druyvesteyn distribution, for example, there would be a large deviation from linearity (Bickerton 1954). With a probability of ionization proportional to the excess energy of the electron over that required for ionization, an expression can be derived for λ in terms of T_e (von Engel and Steenbeck 1932, von Engel 1955, eqn 8.33):

$$\lambda = \frac{600 a p V_i}{\pi^{1/2}} \left(\frac{2kT_e}{m}\right)^{1/2} \left(1 + \frac{2kT_e}{eV_i}\right) \exp\left(-\frac{eV_i}{kT_e}\right) \qquad \ldots\ldots(7)$$

where V_i is the ionization potential, and a is a constant dependent on the gas used. With $2kT_e/eV_i < 1$, and $T_e \gg T_+$, equations (5), (6) and (7) give T_e as a function of Rp. This is shown in figure 3 (labelled ' Schottky ') for helium with $T_+ = 300°\text{K}$.

By equating the number of ions lost to unit length of the tube wall to the number produced in unit length of the column, an estimate of the ratio n_w/n_0 can be made, assuming Schottky's radial distribution. Thus

$$\frac{n_w}{n_0} \simeq \frac{2}{3}\frac{l_+}{R}\frac{T_e}{T_+}. \qquad \ldots\ldots(8)$$

Since in the rare gases $T_e/T_+ \gg 1$, the assumption made in the theory that $n_w/n_0 \ll 1$ becomes invalid even when the mean free path of the ions l_+ is only a small fraction of the tube radius R. The general equation for the radial distribution of electrons (Schottky 1924) remains valid, that is

$$n = n_0 \mathrm{J}_0(xr/R) \qquad \ldots\ldots(9)$$

where $x = (\lambda/D_a)^{1/2}R$ so that the flow density of ions to the tube wall is

$$\frac{j}{e} = \frac{n_w \bar{v}_+}{4} = \frac{n_0 \bar{v}_+}{4} \mathrm{J}_0(x). \qquad \ldots\ldots(10)$$

From (10) and (7), with J_0 and J_1 as Bessel functions of order 0 and 1, the following result is obtained,

$$\frac{4D_a}{Rv_+} = \frac{1}{x}\frac{J_0(x)}{J_1(x)}. \qquad \ldots\ldots(11)$$

With the approximation (6), $D_a = \frac{1}{3}l_+\bar{v}_+T_e/T_+$ and $l_+p = l_{+1}$ equation (11) becomes

$$\frac{4}{3}\frac{l_{+1}}{Rp}\frac{T_e}{T_+} = \frac{1}{x}\frac{J_0(x)}{J_1(x)}. \qquad \ldots\ldots(12)$$

Since x is a function of T_e this equation determines the electron temperature T_e as a function of Rp. Assuming an ion temperature of 300°K, this function is shown in figure 3, labelled 'modified Schottky'.

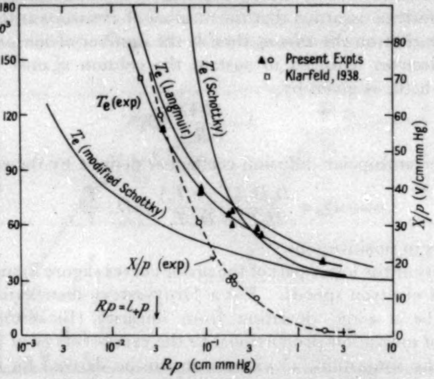


Figure 3. Helium positive column. T_e and X/p plotted against Rp. $V_1 = 24{\cdot}5$ v, $a = 4{\cdot}6 \times 10^{-2}$ (practical units).

The dependence of the rate of ionization λ on the ambipolar diffusion coefficient D_a obtained from (11) is shown in figure 4 for the particular case $R \simeq 2$ cm, $\bar{v}_+ = 1{\cdot}25 \times 10^5$ cm sec^{-1} (corresponding to 300°K in He). On the 'modified Schottky' theory λ becomes almost independent of D_a for high values of D_a, which contrasts with the linear relationship given by the old theory of Schottky.

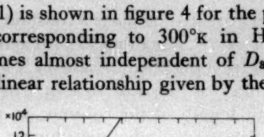


Figure 4. Comparison of theories for (λ, D_a) curves, calculated assuming $R = 1{\cdot}8$ cm, $\bar{v}_+ = 1{\cdot}25 \times 10^5$ cm sec^{-1}. (λ is the number of ionizing collisions per second.)

When the gas pressure is so low that the mean free path of the positive ions is comparable with the radius of the tube, then the ions are able to fall freely to the wall under the influence of the radial electric field where they recombine with the electrons. Theoretical treatment of this case (Tonks and Langmuir 1929) gives for the rate of ionization required to sustain the column against this loss

$$\lambda = \frac{0{\cdot}77}{R}\left(\frac{2kT_\mathrm{e}}{M_+}\right)^{1/2}. \qquad \ldots\ldots(13)$$

The electron temperature for helium calculated from (7) and (13) as a function of Rp is shown in figure 3, labelled 'Langmuir'. The vertical arrow of the figure indicates the value of Rp at which the mean free path of the ions is equal to the tube radius ($l_+ = R$). For lower values of Rp the Langmuir theory should more nearly apply, and for higher values of Rp the modified Schottky theory.

However, the experimental (T_e, Rp) results for zero magnetic field approximate closely to the Langmuir curve over the whole range, although at the higher values of Rp the mean free path for the ions is only a small fraction of the tube radius. It is suggested that this is a consequence of the large radial electric field which results from the relatively high electron temperature, always found in rare gases. This imparts to the ions a motion which is strongly directed towards the tube wall, so that their drift velocity is not a small fraction of their random velocity. Therefore the Boltzmann distribution used in Schottky's theory ($\mu_+/D_+ = e/kT_+$, where μ_+ is the mobility of the ions) is no longer valid. As a result of this rapid carriage of the ions to the wall, the electron temperature required is higher than that given by Schottky's theory. Thus Langmuir's concept of the free fall of the ions to the wall approximates more closely to the result than does Schottky's in which they merely drift radially. The fact that the experimental results lie below the Langmuir curve is due in part to the effect of collisions which retard the radial flow of ions, and in part to two-stage ionization processes, not included in the theory.

5.2. *Power Balance*

If the kinetic energy carried to the walls of the tube is neglected, then from an energy balance in the column a relation can be obtained (von Engel and Steenbeck 1932, von Engel 1955) between the reduced longitudinal electric X/p and the electron temperature (T_e)

$$\frac{X}{p} = \frac{3(3\chi)^{1/2}}{2el_{\mathrm{e}_1}}\,kT_\mathrm{e} \qquad \ldots\ldots(14)$$

where χ is the mean fraction of its random energy that is lost by an electron in a collision with a neutral molecule, and l_{e_1} is the mean free path of an electron at 1 mm Hg gas pressure. From the experimental data of figures 3 and 4 and with the aid of (14), χ as a function of T_e is obtained (figure 8). The rapid rise in χ as the electron temperature is increased shows the increasing importance of inelastic collisions.

5.3. *Effect of the Magnetic Field on the Electron Temperature*

When a magnetic field is applied to the column the diffusion coefficient of the electrons is reduced (Townsend 1915),

$$D = \frac{D_0}{1 + (\omega\tau)^2} \qquad \ldots\ldots(15)$$

where D and D_a are the diffusion coefficients of electrons for a motion perpendicular to $B = B$ and $B = 0$ respectively. Thus the radial electric field required to maintain ambipolar flow is rapidly reduced as B is increased (figure 5 and later discussion). The ions are therefore now allowed to diffuse to the wall with a drift speed small compared with their random velocity as envisaged in the Schottky theory. Since the values of X/p in the column are still relatively high (figure 5) these ions have a mean energy $\bar{\epsilon}_+$ much greater than that corresponding to the gas temperature. An estimate of $\bar{\epsilon}_+$ can be made by applying (14) to the positive ions yielding,

$$\frac{X}{p} = \frac{(3\chi_+)^{1/2}}{el_{+1}}\bar{\epsilon}_+ \qquad \ldots\ldots(16)$$

where χ_+ is the mean fraction of the random energy lost by an ion in a collision with a neutral molecule. For collisions between elastic spheres of equal mass $\chi_+ = 0{\cdot}5$; even with complete charge transfer $\chi_+ \leqslant 1$. With the data of figure 5, and $l_{+1} = 3 \times 10^{-2}$ cm, $\bar{\epsilon}_+$ is found to be of the order 1 ev.

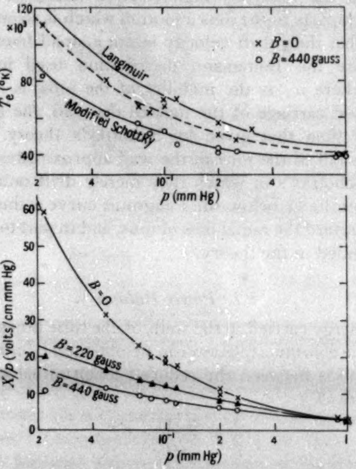


Figure 5. Upper half: T_e as a function of p for $B=0$ and $B=440$ gauss Lower half: X/p plotted against p for $B=0$, $B=220$ and $B=440$ gauss. Helium, $R=1{\cdot}8$ cm.

The positive ions cannot be said to have a temperature corresponding to this energy since their motion is to a large extent in the direction of the field only, instead of random. However due to scattering collisions the ions acquire an effective 'temperature' when diffusing radially which corresponds to a fraction of their mean energy. Taking 0·25 ev as an estimate for this and using the resulting values of T_+ with equation (12), T_e as a function of Rp is obtained (figure 5, 'modified Schottky'). This curve is not very sensitive to the ion temperature. For example, a doubling of T_+ changes T_e by about 3%; the curve may be regarded as accurate in spite of the crude method of estimating T_+.

Figure 5 shows that the magnetic field reduces the electron temperature from the value given by the Langmuir curve to that given by the 'modified Schottky' theory. This agrees well with the theoretical picture given above. Further increase in the magnetic field can only reduce the electron temperature through its influence on the ambipolar diffusion coefficient D_a (equations (6) and (15)). In the range of Rp used in these experiments, however, the rate of ionization λ is very insensitive to changes in D_a (figure 4); only at large B is D_a reduced by orders of magnitude. Thus the reduction in electron temperature through the influence of the magnetic field on the ambipolar diffusion coefficient is here of an altogether smaller order than that due to the influence of B on the radial electric field (transition from Langmuir to modified Schottky theory).

5.4. *Radial Potential Distribution*

Typical results obtained with the movable probe and fixed wall electrode (figure 1) are shown in figure 6. The probe moves into the tube as from the left of the graph. Consequently for probe positions between 3·0 and 3·5 cm from the origin, the probe is in or near the sheath in front of the ring electrode. This explains the strong electric field observed in this region. The apparent asymmetry of the potential distribution is a result of the sheath in front of the movable probe itself. For example, when the probe is 1 cm from the tube wall it is actually floating with respect to the plasma $1+d$ cm from the wall, where d is the thickness of the probe sheath. Because of interference between the two sheaths, no great reliance can be placed on measurements taken at distances smaller than 0·75 cm and greater than 2·75 cm from the origin.

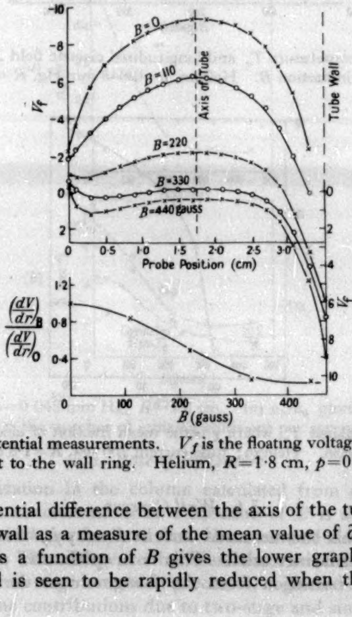


Figure 6. Radial potential measurements. V_f is the floating voltage of the movable probe with respect to the wall ring. Helium, $R=1{\cdot}8$ cm, $p=0{\cdot}12$ mm Hg.

Taking the potential difference between the axis of the tube and the position (d75 cm from the wall as a measure of the mean value of $\partial V/\partial r$, and a plot of $dV/dr)_B/(dV/dr)_0$ as a function of B gives the lower graph of figure 6. The radial electric field is seen to be rapidly reduced when the magnetic field is increased.

5.5. *Longitudinal Electric Field*

The relation between X_0, the longitudinal electric field for $B=0$ and X_B that in the presence of a field B, can be obtained from an extension of (14):

$$X_B/X_0=(T_{eB}/T_{e0})(\chi_B/\chi)^{1/2}. \qquad \ldots\ldots(17)$$

X_B, as a function of B, obtained in this way, is compared in figure 7 with the experimentally determined values. In using equation (17) the measured values of T_e (figure 5) were taken, and χ_B derived from figure 8. The good agreement obtained serves to confirm the measurements of electron temperature, since equation (14) is essentially a relation between T_e and X/p. It also shows that no new processes become active in the discharge when a magnetic field is applied, since $\chi=f(T_e)$ is independent of B and p. The character of this function agrees with observations of χ in other gases (von Engel 1955, p. 108).

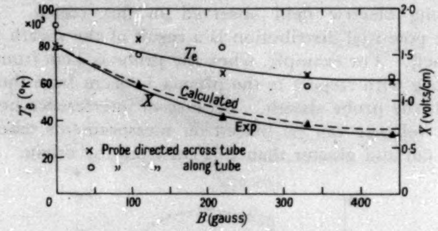


Figure 7. Electron temperature T_e and longitudinal electric field X as a function of the magnetic induction B. Helium, $p=0{\cdot}048$ mm Hg, $R=1{\cdot}8$ cm.

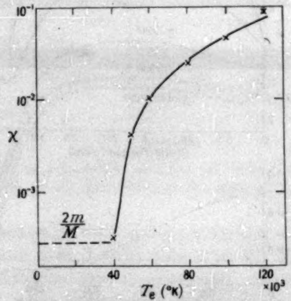


Figure 8. χ plotted against T_e, where χ is the mean fraction of electron energy lost per collision. Helium; tube current 100 mA, $R=1{\cdot}8$ cm.

5.6. *Wall Current*

Figure 9 shows that the measured current density j_+ of ions to the wall falls as the magnetic field is increased. This is in qualitative agreement with the theory indicating a transition from the Langmuir to the Schottky concept (figure 5).

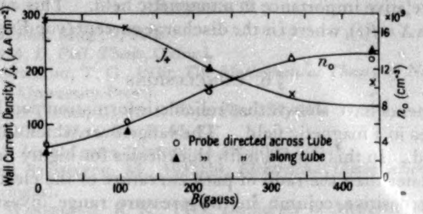


Figure 9. Axial electron concentration n_0 and positive ion density j_+ at the wall as a function of the magnetic induction B. Helium, $p=0{\cdot}048$ mm Hg, $R=1{\cdot}8$cm.

5.7. *Axial Electron Density*

The axial electron density n_0 rises rapidly with increasing magnetic field (figure 9). This is partly accounted for by the reduction of the ratio n_w/n_0 as B is increased (figure 10). In order to maintain the same number of electrons per cm length of column, n_0 must rise if the electron concentration at the tube wall n_w is reduced. Finally the reduction of X/p in the column and hence of the longitudinal drift speed of electrons means that to maintain a constant current the number of electrons per unit length must increase as the magnetic field is raised.

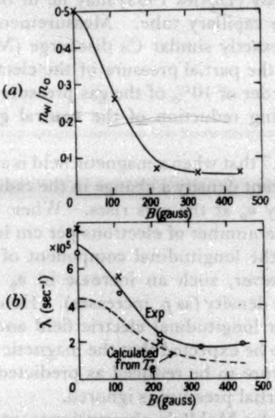


Figure 10. Helium, $p=0{\cdot}048$ mm Hg, $R=1{\cdot}8$ cm. (*a*) n_w/n_0 plotted against B. (*b*) λ plotted against B. (λ is the number of ionizing collisions per second and per electron.)

5.8. *Rate of Ionization*

The rate of ionization in the column calculated from the measured wall current and electron concentrations as a function of B is shown in figure 10. The broken line shows λ as calculated from the measured electron temperatures using equation (7). The difference between the two curves represents the contribution due to two-stage ionization processes, neglected in (7). Within the experimental error the contributions due to two-stage and single stage processes have the same relative importance in a magnetic field. This agrees with the fact that in rare gases $X = f(i)$, where i is the discharge current (von Engel and Steenbeck 1934).

§ 6. CONCLUSIONS

The experiments have shown that reliable information can be obtained from the use of probes in a magnetic field. The range over which this is valid has not been determined. In this work sheath thicknesses for highly negative potentials were always greater than the radii of path curvature of the electrons.

The helium positive column in the pressure range investigated has been found, in the absence of a magnetic field, to behave in closer agreement with the theoretical predictions of Langmuir for the case $l_+ > R$, than those of Schottky for $l_+ < R$. This is because the positive ions acquire from the large radial electric field a strongly directed motion more akin to that in vacuum (Tonks and Langmuir), than to a slow radial drift accompanied by random motion (Schottky).

When a longitudinal magnetic field is applied to the column the electron temperature has been found in all cases to be reduced. Other discharge parameters show dependences on B at least in qualitative agreement with theory. With the current density used ($\simeq 10$ mA cm^{-2}) no evidence has been found for an increase in the diffusion coefficient of electrons across a magnetic field in the positive column due to fluctuations of electron density or to oscillations.

It is suggested that the reason for the discrepancy between theory and experiment found previously (Davies 1953) may lie in the high current density of 5 A cm^{-2} then used in a capillary tube. Measurements of the electron concentration made with an exactly similar Cs discharge (Mohler 1938) show that since $T_e \simeq 4000 \ldots 5000$°K the partial pressure of the electrons ($p_e = n_0 k T_e$) at the axis of the tube is of the order of 10% of the gas pressure at $p = 8 \times 10^{-2}$ mm Hg. This causes a corresponding reduction of the neutral gas density on the axis (Steenbeck 1939).

It has been shown in § 5.7 that when a magnetic field is applied to a low pressure positive column at low current density a change in the radial electron distribution takes place and the density n_0 at the axis rises. When the current is low and constant, the increase in the number of electrons per cm length of tube is accompanied by a reduction in the longitudinal component of the electric field. At high current density, however, such an increase in n_0 would cause a drastic reduction in the neutral gas density (as p_e increases). Hence to produce electrons at the former rate a higher longitudinal electric field and electron temperature are required. Thus it is to be expected that the magnetic field will not cause the field and electron temperature to be reduced as predicted by a theory in which the effect of the electron partial pressure is ignored.

This view is supported by Mohler's observations; as the current density is increased at constant pressure, both X and T_e rise; the most probable cause of the rise in T_e being the increase in the partial pressure p_e. There may even be a tendency for T_e to become dependent on radial position since the magnetic field severely restricts radial diffusion. In this way an observed rise in T_e might be explicable.

ACKNOWLEDGMENTS

The authors thank Lord Cherwell for the facilities of the Clarendon Laboratory and Professor K. G. Emeléus for valuable suggestions.

REFERENCES

BAILEY, V. A., 1930, *Phil. Mag.*, **9**, 560.

BICKERTON, R. J., 1954, *D. Phil. Thesis*, Oxford.

CHAPMAN, S., and COWLING, T. G., 1939, *The Mathematical Theory of Non-Uniform Gases* (Cambridge : University Press).

CUMMINGS, C. S., and TONKS, L., 1941, *Phys. Rev.*, **59**, 514.

DAVIES, L. W., 1953, *Proc. Phys. Soc.* B, **66**, 33.

VON ENGEL, A., 1955, *Ionized gases*, (Oxford : Clarendon Press).

VON ENGEL, A., and STEENBECK, M., 1932/34, *Elektrische Gasentladungen*, vol. 1 and 2 (Berlin : Springer).

GUTHRIE, A., and WAKERLING, R. K., 1949, *Discharges in Magnetic Fields* (New York : McGraw-Hill), p. 197.

LANGMUIR, I., and MOTT-SMITH, H., 1924, *Gen. Elect. Rev.*, **27**, 449.

MOHLER, F. L., 1938, *J. Res. Nat. Bur. Stand.*, **21**, 873.

SCHOTTKY, W., 1924, *Phys. Z.*, **25**, 635.

SIMON, A., 1955, *Phys. Rev.*, **98**, 317.

STEENBECK, M., 1939, *Wiss. Veröff. Siemens*, **18**/3, 94.

SPIWAK, G., and REICHRUDEL, E., 1938, *J. Tech. Phys. U.S.S.R.*, **5**, 715.

TONKS, L., 1939, *Phys. Rev.*, **59**, 514.

TONKS, L., and LANGMUIR, I., 1929, *Phys. Rev.*, **34**, 876.

TOWNSEND, J. S., 1915, *Electricity in Gases* (Oxford : University Press).

RESEARCH NOTES

Heat Capacity of Silver below 4·2°K

By J. A. Rayne
Division of Physics, Commonwealth Scientific and Industrial Research Organization, Sydney

MS. received 18th November 1955

At liquid helium temperatures, the heat capacity of a normal metal may be expressed in the form

$$C=\gamma T+A(T/\theta)^3$$

the first and second terms representing the electronic and lattice specific heats respectively. When fitted to this equation, the low-temperature heat capacity data for silver gives the following values of γ and θ:

Observer	γ (Joule mole^{-1} deg^{-2}) $\times 10^3$	θ (deg K)
Keesom and Kok (1932, 1933/34)	$0{\cdot}67 \pm 0{\cdot}02$†	230 ± 2†
Keesom and Pearlman (1952, 1955)	$0{\cdot}66 \pm 0{\cdot}01$†	229 ± 1†
Corak, Garfunkel, Satterthwaite and Wexler (1955)	$0{\cdot}610 \pm 0{\cdot}005$	$225{\cdot}3 \pm 0{\cdot}2$

† Errors estimated by present author from plots of C/T against T^2.

The latter two sets of results pertain to specimens of 99·999% and 99·98% purities respectively, and in both cases corrections were made for deviations from the 1948 temperature scale (Erickson and Roberts 1954). Since the deviations between the measurements are considerably larger than the estimated errors, a fresh determination of the heat capacity of silver below 4·2°K was undertaken.

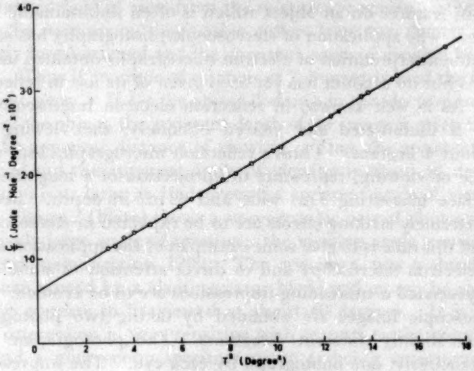
For the present experiment, a silver specimen of 99·99% purity and mass 199·8 grammes was used. Measurements were made on a calorimeter using a mechanical heat switch rather than exchange gas, in order to eliminate the possibility of spurious effects due to desorption of gas during heating periods. The design is similar to that described by Ramanathan and Srinivasan (1955), except that provision is made for continuous evacuation of the working space. A 10 ohm, Allen–Bradley carbon resistance is used as a thermometer. Calibration of the thermometer was effected, at the conclusion of the experiment, against the vapour pressure of liquid helium using the Erickson and Roberts corrections to the 1948 temperature scale. To ensure equilibrium between the thermometer and the helium bath, exchange gas was admitted to the working space during calibration rather than relying on physical contact between the specimen and the bottom of the vacuum can.

The results of the experiment are presented in a graph of C/T against T^2 (see figure). The straight line represents a least squares fit to the experimental points, the maximum deviation of any point from the line being about 2%.

The corresponding values of γ and θ are

$$\gamma = 0{\cdot}613 \pm 0{\cdot}005 \times 10^{-3}\,\text{joule mole}^{-1}\,\text{deg}^{-2}$$
$$\theta = 226{\cdot}2 \pm 0{\cdot}3\,\text{deg K.}$$

The uncertainties quoted in these figures are the random errors associated with the dispersion of the points about the straight line of best fit in the plot of C/T against T^2. Exchange gas has been eliminated as a possible cause of inaccuracy and a careful analysis gives an upper limit of 0·5% for the remaining sources of systematic error. As the results of the present work and those of Corak *et al.* (1955) agree within the resulting combined error even though the specimens came from different sources, it can reasonably be concluded that the outstanding disagreement with the other work is due not to differences in the physical states of the specimens but to undetected systematic errors in the older experiments. One likely cause of this error arises from the use of exchange gas. This tends to give too large a value of heat capacity, the effect being more pronounced the lower the temperature. The corresponding values of γ and θ would then be too large.



(C/T, T^2) for silver.

ACKNOWLEDGMENTS

The author would like to express his thanks to Mr. W. R. G. Kemp for his assistance during the course of this work. Thanks are also due to Mr. J. W. Smyth for providing the liquid helium used during the experiments.

REFERENCES

CORAK, W. S., GARFUNKEL, M. P., SATTERTHWAITE, C. B., and WEXLER, A., 1955, *Phys. Rev.*, **98,** 1699.

ERICKSON, R. A., and ROBERTS, L. D., 1954, *Phys. Rev.*, **93,** 957.

KEESOM, P. H., and PEARLMAN, N., 1952, *Phys. Rev.*, **88,** 140; 1955, *Ibid.*, **98,** 548.

KEESOM, W. H., and KOK, J. A., 1932, *Leiden Comm.*, No. 219d; 1933/1934, *Ibid.*, No. 232d.

RAMANATHAN, K. G., and SRINIVASAN, T. M., 1955, *Phil. Mag.*, **46,** 338.

Stereoscopic Reflection Electron Microscopy

By D. E. BRADLEY, J. S. HALLIDAY and W. HIRST
Research Laboratory, Associated Electrical Industries, Ltd., Aldermaston, Berks.

MS. received 12*th December* 1955

STEREOSCOPIC photographs provide information about the positions in depth of features on an object which is often unobtainable from a single picture. The application of stereoscopic photography has already proved invaluable in the interpretation of electron micrographs obtained using transmission specimens but no account has yet been given of its use in reflection electron microscopy. As is well known, in reflection electron microscopy the surface of an object is illuminated and viewed obliquely, the viewing angle being commonly about 4 degrees. Thus a reflection micrograph, 3 in. square, taken with this angle of viewing, represents the projection of a magnified area of the specimen surface measuring 3 in. wide and 43 in. in depth. Because of this great depth extremely striking effects are to be expected in stereoscopic viewing. The purpose of this note is to give some examples of the application of stereoscopy in reflection electron microscopy and to direct attention to some effects which need to be appreciated if misleading impressions are to be avoided.

The stereoscopic images are obtained by taking two photographs of the specimen from slightly different directions. The photographic pair is then viewed simultaneously, one photograph by each eye. The impression of depth obtained by stereoscopic viewing with the eyes arises because the angular difference in position of distant objects is smaller than that of near objects. Thus, if the pair of micrographs is to appear to the eyes as a magnified image of the specimen in the original perspective, the angular displacements between points in the two micrographs must vary in the same manner with respect to the eyes as do the corresponding points in the specimen with respect to the objective lens of the electron microscope. As a typical example, one may consider the instrument to be used at a lateral magnification of 1500, the minimum usable object distance in the electron microscope being 0·8 in. Then if the angle of viewing is 4 degrees the area of the object which is observed comprises a rectangle 0·002 in. in width and 0·029 in. in depth, and the ratio of the depth of the field of view (which is greater than the depth of focus, see Haine and Hirst (1953)) to the distance from the objective lens is 0·029/0·8. If the photographic images are to give the correct stereoscopic impression the ratio of the magnified depth of the object to the apparent viewing distance must also be equal to 0·029/0·8. Thus for the images to have an apparent depth of 43 in. they must seem to be $43 \times 0{\cdot}8/0{\cdot}029$ in. away, that is 1190 in. For an interocular distance of 3 in., this could be achieved by rotating the specimen through an angle of $1{\cdot}26 \times 10^{-3}$ radian, which is just over $\frac{1}{15}$ degree, or by translating it by 0·001 in. If stereoscopic images of object points further from the objective lens are required then an even smaller degree of rotation or translation will be needed. Unfortunately with our present specimen manipulator it is not possible to control the extent of rotation or translation to such a high degree of accuracy. However, attempts have been made to achieve this small rotation by chance and figure 1

(Plate) shows a stereographic pair of photographs obtained in this way. The specimen consisted of graphite flakes left on a polished brass surface after drawing a pencil line across it, the lateral magnification was 1200 and the angle of viewing 4 degrees.

Stereoscopic viewing still gives a three-dimensional effect if a greater rotation is used and for some purposes, such as the determination of the relative positions of features little separated in depth, it may be necessary to use large angles of rotation to obtain sufficient stereoscopic effect. (Translation of the specimen cannot be used to obtain this effect since the large movements required would remove the desired field of view from the microscope screen.) On viewing such photographs the stereoscopic impression suffers from two defects, the specimen appears greatly foreshortened and the apparent angle of viewing becomes greatly exaggerated. Thus if an angle of rotation of 2 degrees is used the image appears to be only 86 in. away. If l is the object distance in the electron microscope and Δl the depth of the object, the apparent depth of the image is given by $\Delta L = L\Delta l/l$, where L is the apparent distance of viewing. Thus the apparent depth of the image is 3·1 in. and the apparent angle of viewing is about 44 degrees. If the angle of rotation is as large as 10 degrees the apparent angle of viewing becomes 80 degrees. Figure 2 (Plate) shows a stereographic pair of photographs obtained using a relatively large angle of rotation; the viewing angle was again 4 degrees and the lateral magnification 1200. The specimen was a duralumin surface which had been scored by a slanting razor blade and as can be seen the stereographic image appears to be viewed at about 40 degrees. It is clear that the stereoscopic impression is very sensitive with respect to the angular rotation of the object and a microscope specimen stage is being constructed which will enable precise rotational and translational control to be obtained.

In order to obtain the correct stereoscopic impression of the shadows cast by roughnesses on the object surface, the source of illumination should be rotated through the same angle as the object. However, when obtaining the photographic pairs shown in figures 1 and 2 the source of illumination remained stationary. In consequence of this the shadows appear to lie vertically below the specimens, similar to the images one observes on a water surface. This effect would be avoided if the stereoscopic images were obtained by translation of the specimen.

ACKNOWLEDGMENTS

The authors wish to express their thanks to Mr. M. E. Haine for the use of the electron microscope, and to Dr. T. E. Allibone for permission to publish this note.

REFERENCE

HAINE, M. E., and HIRST, W., 1953, *Brit. J. Appl. Phys.*, **4**, 239.

LETTERS TO THE EDITOR

On High Temperature Expansions in the Heisenberg Theory of Ferromagnetism

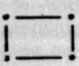
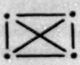
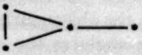
In a recent article in this journal Rushbrooke and Wood (1955) discuss the calculation of the high temperature susceptibility for the Heisenberg model of a ferromagnetic. One of the basic problems arising is the calculation of averages of spin operators for various cluster sums arising in the expansion. Rushbrooke and Wood point out that the calculations are very involved, and that it would be rather desirable to provide a check on the calculations.

We have been working independently on the same problem, and initially calculated the averages in the same way as Rushbrooke and Wood, i.e. by using the commutation relations of the spin operators and reducing the various clusters. We found, however, that the method rapidly became exceedingly complicated, and we therefore devised an alternative procedure which proved appreciably simpler, and could also act as a check. This alternative method calculates the averages of the permutation operators P_{ij}, rather than the spin operators $\boldsymbol{\sigma}_i \, . \, \boldsymbol{\sigma}_j$, using the Dirac relation

$$P_{ij} = \tfrac{1}{2}(1 + \boldsymbol{\sigma}_i \, . \, \boldsymbol{\sigma}_j).$$

It is much easier to determine the resultant permutation which is the product of a number of permutation operators; if this resultant permutation is expressed in the form of cycles, each of these cycles can be treated independently, and the averages readily determined.

As an additional check we determined the exact partition functions for systems consisting of a finite number of spins such as

When these are expanded at high temperatures each term consists of a sum of cluster averages, and we are provided with an independent check on the correctness of the calculation of the averages.

Our results check precisely with those of Rushbrooke and Wood as far as the latter are tabulated, but we have also been able to calculate an additional term as follows :

$a_6 = -133/480$ (linear chain), $367/720$ (simple quadratic), $8143/480$ (triangular), $33013/480$ (simple cubic), $16993/24$ (body centred cubic), $914601/80$ (face centred cubic).

Further details of the method described above, and of other related calculations, together with the physical significance of the results will be given in a subsequent paper.

† Wheatstone Laboratory,
King's College, London.
‡ Laboratoire d'Électrostatique et de Physique du Metal,
Grenoble.
† C. Domb.
‡ M. F. Sykes.

30th December 1955.

Rushbrooke, G. S., and Wood, P. J., 1955, *Proc. Phys. Soc.* A, **68**, 1161.

Time Dependent Changes of Surface Lifetime in Germanium in the presence of Electric Fields

In a recent paper, Henisch and Reynolds (1954) have reported on the modulation of the surface recombination velocity of germanium by an electrostatic field applied normally to the surface of the specimen. The experiments were carried out with mica sheets between the germanium and the electrodes used for applying the field. These measurements are being extended in a system in which mica spacers are not employed. The results now show that the changes of surface lifetime are time dependent. In order to study this variation the change in conductance of an illuminated specimen upon application of an electrostatic field is measured by a Wheatstone bridge, the out of balance signal being amplified and displayed on a recording milliammeter. The changes measured are at least an order of magnitude larger than those arising from modulation of surface conductance (Low 1954).

Provided that the dimensions of the specimen are such that the thickness is not large compared with the diffusion length for carriers, and that there is no appreciable sweeping-out, such a change may be related to a change in filament lifetime. It can be shown that for small signals the conductance of the specimen under illumination and the application of a field may be written

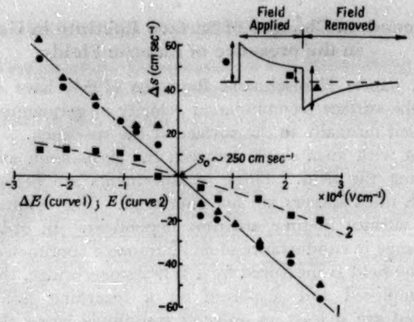
$$\frac{A}{l}(n\mu_n+p\mu_p)e+\frac{N\tau_0}{l^2}(\mu_n+\mu_p)e+\frac{N\Delta\tau}{l^2}(\mu_n+\mu_p)e$$

where n, p, μ_n, μ_p, are the concentrations and mobilities of electrons and holes, A and l the cross-sectional area and length of the specimen, N the number of minority carriers injected optically per second into the specimen, and τ_0, $\Delta\tau$ the filament lifetime with zero applied field and the change in lifetime upon application of a field. The first term is the thermal equilibrium conductance. The second gives the increase in conductance due to illumination in the absence of an applied field, and the third the effect of the applied field on this increase. The ratio of the last two terms gives $\Delta\tau/\tau_0$, where τ_0 may be measured independently. With a knowledge of the geometry of the specimen the change in surface recombination velocity Δs may be calculated from $\Delta\tau$.

The specimens used were of dimensions approximately 2 cm × 1 cm × 0·1 cm, with non-injecting soldered end contacts. The modulating field was applied normally to the two large faces by two electrodes at the same potential, each spaced from the germanium by two glass rods of about 0·025 cm diameter.

In each case it was observed that the application of a field produced a sudden conductance change which decayed to a small value in a time of the order of 200 seconds. Removal of the field produced a behaviour similar but opposite in sign. Owing to the inevitable presence of drift in the apparatus the assumed steady state value could not be measured with the same degree of accuracy as the initial changes. However, systematic differences were obtained between the steady-state conditions for various applied fields, and taking the zero field value as a datum, results were obtained of the type shown in the figure. In the steady-state condition, sudden reversal of the field caused a swing of twice that due to merely switching it off.

All measurements have been made after etching the specimen in CP4 and so far have been conducted in dry air, mainly at atmospheric pressure. Four



Curve 1 : The initial change in *s* as a function of ΔE, a sudden change in applied field.
Curve 2 : The steady state change in *s* as a function of the applied field *E*. Positive values of *E* correspond to positive induced charge on the germanium surface.
Inset : The time dependence of the conductance on application and removal of a field.

crystals, of which three were n-type, of resistivities 10, 12 and 45 ohm cm, and one p-type of 10 ohm cm, yielded a dependence of lifetime on applied field of which the figure is typical. However, one sample gave a dependence of *s* reversed in sign from that shown in the figure. This sample was n-type and had a resistivity of 20 ohm cm. Its anomalous behaviour was repeatedly observed after many successive etchings.

Acknowledgment is made to the Admiralty for permission to publish this letter.

Physics Research Laboratory,
University of Reading.
12th December 1955.

J. D. Nixon.
P. C. Banbury.

Henisch, H. K., and Reynolds, W. N., 1954, *Proc. Phys. Soc.* B, **68**, 353.
Low, G. G. E., 1954, *Proc. Phys. Soc.* B, **68**, 10.

The Absorption of 39 kMc/s (39 Gc/s) Radiation in Germanium

§ 1. Theoretical

The equation

$$E = \exp(i\omega t - iLz)$$

describes a transverse electromagnetic wave propagated in the *z* direction through a medium which is free from space charge and has a propagation constant *L*. If the attenuation is due to the motion of free charge carriers it can be shown, from consideration of the equations of motion of the electrons and holes, together with the field equations, that

$$L^2c^2 = \omega^2 - \frac{q^2N_1}{\epsilon m_1}\left(\frac{1}{1-iq/\omega m_1\mu_1}\right) - \frac{q^2N_2}{\epsilon m_2}\left(\frac{1}{1-iq/\omega m_2\mu_2}\right) \quad \ldots\ldots(1)$$

in rationalized m.k.s. units. In equation (1) N_1 and N_2 are the densities of electrons and holes respectively, m_1 and m_2 their masses and μ_1 and μ_2 their mobilities; q is the electronic charge, c the velocity of light in the medium of permittivity ϵ.

When applied to a metal, equation (1) may be simplified to give the classical expression for the attenuation constant. For a semiconductor of not too high conductivity we may assume that $\epsilon\omega \gg q\mu_1 N_1$ or $q\mu_2 N_2$. Two cases then arise:

(1) For infra-red radiation we may assume that $\omega \gg q/m_1\mu_1$ or $q/m_2\mu_2$. The absorption constant, equal to twice the imaginary part of L, is then given by

$$K=\lambda^2\left[\left(\frac{\mu_0}{\epsilon_0}\right)^{1/2}\frac{q^3}{4\pi^2 n c_0{}^2}\right]\left[\frac{N_1}{m_1{}^2\mu_1}+\frac{N_2}{m_2{}^2\mu_2}\right] \qquad \ldots\ldots(2)$$

where $n=\sqrt{\epsilon}$ is the refractive index.

It has been known for some years, however, that equation (2) indicates an absorption constant about 1000 times smaller than that observed in germanium and other semiconductors (Becker and Fan 1951, p.132, Gibson 1952, 1953) when the electronic mass is used for m_1 and m_2.

(2) For microwave or radio frequencies we may assume that $\omega \ll q/m_1\mu_1$ or $q/m_2\mu_2$. The absorption constant is then given by

$$K=\frac{n}{c_0}\left[\frac{\sigma}{\epsilon}\right]=1635\frac{\sigma}{n}\ \text{dB per metre} \qquad \ldots\ldots(3)$$

where $\sigma=q(\mu_1 N_1+\mu_2 N_2)$ is the conductivity for direct current.

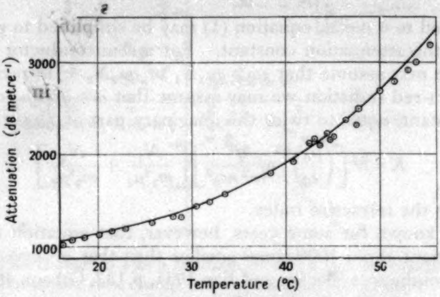
§ 2. EXPERIMENTAL

Equation (3) is of basic importance in the operation of germanium microwave modulators of the type described by Gunn and Hogarth (1955). In view of the marked discrepancy between theory (equation (2)) and experiment for semiconductors in the infra-red region of the spectrum it appeared to be desirable to check the validity of equation (3) by direct measurement at 8·6 mm wavelength. As the refractive index of germanium at 1·25 cm wavelength is known to be about 4·05 (Goldey and Brown 1955) a sample thick compared with the wavelength, actually 1·91 cm thick, was prepared. The sample was cut and ground with great care to fit a section of electroformed waveguide, the gap between sample and guide nowhere exceeding 0·0002 inch. The room temperature resistivity of the sample was 40 ohm cm, thus ensuring the validity of the conditions assumed above.

The absorption in the specimen was measured from room temperature to 55°C at which temperature the transmitted signal became too small for accurate measurement (sample attenuation, 65 dB). The conductivity of the sample over the same temperature range was measured in a separate experiment.

The results of this experiment are given in the figure. The curve is the theoretical absorption constant, derived from equation (3), using the experimental data on conductivity and assuming a value of the refractive index n equal to 4·1. The experimental points indicate the observed absorption constant. It will be seen that the agreement is excellent over most of the temperature range. Hence it seems likely that the high infra-red absorption is associated with the small effective mass of the carriers (Goldey and Brown 1955), a quantity which does not appear in equation (3).

The experimental results described are not in agreement with those of Klinger (1953). According to Klinger, the radio-frequency absorption is less than that indicated by the d.c. conductivity, particularly at elevated temperatures.



Experimental points : observed attenuation.
Curve calculated from 1635 σ/n taking σ experimental, $n=4{\cdot}1$.

Klinger's results indicate a very large value of the carrier relaxation time τ namely 4×10^{-11} second. The value of τ is related to the effective masses by the equations $\tau_1=\mu_1 m_1/q$ and $\tau_2=\mu_2 m_2/q$ and hence the above value of τ indicates effective masses about ten times greater than the electronic mass. As recent work, referred to above, indicates an effective mass appreciably less than the electronic mass, Klinger's results appear to be difficult to interpret. Our results indicate that the condition assumed in deriving equation (3) is valid and hence

$$\tau_1, \tau_2 \ll 4\times10^{-12}\text{ second}; \quad m_1 \ll 2m; \quad m_2 \ll 4m.$$

I am indebted to various colleagues at the Radar Research Establishment for the major part of the above work, namely Dr. C. A. Hogarth and Mr. J. B. Gunn for preparing the sample, Mr. J. B. Arthur for the absorption measurements and Dr. G. G. Macfarlane for bringing equation (1) to my notice. This note is published by permission of the Controller, H.M. Stationery Office.

Radar Research Establishment,
Malvern, Worcs.
22nd December 1955.

A. F. Gibson.

Becker, M., and Fan, H. Y., 1951, *Semi-conducting Materials*, Ed. H. K. Henisch (London : Butterworths Scientific Publications), p. 132.
Gibson, A. F., 1952, *Proc. Phys. Soc.* B, **65**, 378; 1953, *Ibid.*, **66**, 588.
Goldey, J. M., and Brown, S. C., 1955, *Phys. Rev.*, **98**, 1761.
Gunn, J. B., and Hogarth, C. A., 1955, *J. Appl. Phys.*, **26**, 353.
Klinger, Y., 1953, *Phys. Rev.*, **92**, 509.

The Effect of Field Emission on the Behaviour of Semiconductor Contacts

In a recent paper, Sillars (1955, to be referred to as I) derived the voltage-current characteristics of semiconductor point contacts in the absence of barrier effects and assuming field emission across the narrow gap in the vicinity of the mechanical contact. The analysis was carried out for gap widths g which are proportional to a power of the radial displacement r from the centre of the contact and under the assumption of a constant field F_0 across the gap in the region of field emission. In this note the analysis is extended to include arbitrary gap widths and fields which may be a function of r. The latter case might arise if there is a variable work function χ at the gap surfaces since the predominant factor determining the field F_0 is the term

$$\exp\left[\frac{4}{3}\left(\frac{8m\pi^2}{h^2}\right)^{1/2}\frac{\chi^{3/2}}{F_0}\right]$$

which occurs in the expression for the field emission current. Thus $F_0 \propto \chi^{3/2}$, the constant of proportionality depending on the range of current densities (cf. Discussion in I which justifies the assumption that F_0 is approximately independent of current density).

Introducing the dimensionless quantities $\lambda = r/c$; $v = V_S/V_0$; $i_s = \frac{1}{2}\pi I_S c/V_0\sigma$; $k = J/4\sigma c V_0$ where I_S and V_S are the current density and voltage at any point r, J and V_0 are the total current and applied voltage, σ is the conductivity and c is an arbitrary length to be specified, equations (8), (9), (10) and (11) of I can be written as

$$1 - v = \frac{2}{\pi}\int_0^\lambda f(x)\cos^{-1}\frac{x}{\lambda}\,dx \quad \ldots\ldots(1) \qquad 1 = \int_0^\infty f(x)dx \quad \ldots\ldots(3)$$

$$i_S = \int_\lambda^\infty \frac{f(x)dx}{(x^2-\lambda^2)^{1/2}} \quad \ldots\ldots(2)$$

[illegible] $= \int_0^\infty xf(x)dx. \quad \ldots\ldots(4)$

In order to obtain a direct relation between k and v, the function $p(x)$ will be introduced where

$$p(x) = 1 - \int_0^x f(y)dy = \int_x^\infty f(y)dy \qquad \ldots\ldots(5)$$

so that

$$p(\infty) = 0 \text{ and } p(0) = 1. \qquad \ldots\ldots(6)$$

Substituting into equation (4) and (1)

$$k = \int_0^\infty p(x)dx \text{ if } \lim_{x\to\infty} xp(x) = 0 \qquad \ldots\ldots(7)$$

(if the upper limit of all the integrals is finite, as will be the case in the problem treated below, equation (7) is always satisfied) and

$$v = \frac{2}{\pi}\int_0^\lambda \frac{p\,dx}{(\lambda^2 - x^2)^{1/2}}. \qquad \ldots\ldots(8)$$

This can be reduced to Abel's integral equation by a suitable transformation and solved for $p(x)$, viz.

$$p(x) = \frac{d}{dx}\int_0^x \frac{\lambda}{(x^2-\lambda^2)^{1/2}}\,v(\lambda)d\lambda. \qquad \ldots\ldots(9)$$

Substituting into equation (7)

$$k=\lim_{x\to\infty}\int_0^x \frac{\lambda}{(x^2-\lambda^2)^{1/2}}\, v(\lambda)d\lambda. \qquad \ldots\ldots(10)$$

If the upper limit of the integrals is finite, say $i_s(\lambda)=0$, if $\lambda>\lambda_c$ (i.e. $f(\lambda)$, $p(\lambda)=0$ if $\lambda>\lambda_c$) then

$$k=\int_0^{\lambda_c} \frac{\lambda}{(\lambda_c^{\,2}-\lambda^2)^{1/2}}\, v(\lambda)d\lambda \qquad \ldots\ldots(11)$$

and equation (6) will be replaced by

$$p(\lambda_c)=\left[\frac{d}{dx}\int_0^x \frac{\lambda}{(x^2-\lambda^2)^{1/2}}\, v(\lambda)d\lambda\right]_{x=\lambda_c}=0 \qquad \ldots\ldots(12)$$

using equation (9).

Since c has been left undetermined, we can set $\lambda_c=1$. Also simplifying equations (11) and (12) by setting $\lambda=\lambda_c\sin\theta$ and $\lambda=x\sin\theta$ respectively gives the two equations

$$k=\int_0^{\pi/2} v\,(\sin\theta)\,\sin\theta\,d\theta=-\int_0^{\pi/2} v'\,(\sin\theta)\sin^2\theta\,d\theta \qquad \ldots\ldots(13)$$

where

$$r=c\sin\theta. \qquad \ldots\ldots(14)$$

As in I it will now be supposed that the voltage drop across the gap is given by

$$V_0-V_S=F_0(r)\,g(r) \qquad \ldots\ldots(15)$$

where F_0 and g have to be specified.

In the reduced variables equation (15) becomes

$$1-v=F_0 g(c\sin\theta)/V_0. \qquad \ldots\ldots(16)$$

If we introduce the integral

$$U(c)=\int_0^{\pi/2} F_0 g(c\sin\theta)\sin\theta\,d\theta \qquad \ldots\ldots(17)$$

the two equations (13) may be reduced to

$$V_0=U(c)+cU'(c) \qquad \ldots\ldots(18)$$

and

$$k=\frac{cU'(c)}{U(c)+cU'(c)} \quad \text{or} \quad J=4\pi c^2U'(c). \qquad \ldots\ldots(19)$$

Evaluation of the integral $U(c)$ and elimination of c between equations (18) and (19) would give the relation between J and V_0 thus providing a general and straightforward method of calculating currents and voltages, whatever the contact geometry or the variation of work function in the neighbourhood of the contact may be.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. R. W. Sillars for helpful discussions of the problem, Dr. Willis Jackson, Director of Research and Education, and Mr. B. G. Churcher, Manager of the Research Department, Metropolitan–Vickers Electrical Co. Ltd., for permission to publish this paper.

Research Department,
Metropolitan-Vickers Electrical Co. Ltd.,
Trafford Park, Manchester.
R. STRATTON.

19th December 1955.

R. W. SILLARS, 1955, *Proc. Phys. Soc.* B, **68**, 881.

REVIEWS OF BOOKS

Operational Calculus based on the two-sided Laplace Integral, by B. VAN DER POL and H. BREMMER. 2nd Edn. Pp. xiv + 415. (Cambridge: University Press, 1955.) 60*s.*

Physicists interested in the application of modern operational methods to the study of physical problems will already be familiar with the first edition of this book which appeared in 1950. In this work the authors base an operational calculus upon the properties of the bilateral Laplace transform

$$f(p)=\int_{-\infty}^{\infty} e^{-pt}h(t)dt,$$

where it is assumed that p and hence $h(t)$ are such as to ensure the existence of the integral. The theory of such transforms is developed in the first half of the book and the second half is devoted to the solution of ordinary differential equations with constant and variable coefficients, of systems of simultaneous ordinary differential equations and of certain partial differential equations. There is a large number of examples drawn from a wide field in both physics and pure mathematics. The whole work is very well summarized in a 'Grammar' which is a collection of general formulae and rules of manipulation for the operational calculus, and a 'Dictionary' containing classified transform pairs.

The first edition of this book was distinguished by the high quality of the writing and the novel approach adopted by the authors. The concepts of Fourier theory are well explained and illustrated but the proofs of the basic theorems are omitted, reference being made to the classic treatises of Doetsch and Widder. This is probably the most effective method of treating the subject for readers whose main interest lies not in the theory but in its application to physical problems and there is no doubt that the authors have carried out their plan with signal success.

The new edition differs only slightly from the first. A number of corrections and improvements have been made at various places but only three sections of essentially new material have been added. Rules for the treatment of correlation functions have been added at the end of the chapter on 'Elementary Rules' (Chapter IV). The remarks on the Dirac delta function have been supplemented by a very brief paragraph which does no more than mention Schwarz's theory of distributions. A brief note on the operational interpretation of the Wiener–Hopf technique has been added to the chapter on integral equations (Chapter XIV). These additions are much too brief to add substantially to the stature of the work; this was already great and there is no doubt that this book will become a standard work on operational methods. I. N. SNEDDON.

Journées de mécanique des fluides. Pp. 331. (Paris: Publications Scientifiques et Techniques de l'Air, 1955.) 3000*fr.*

The first part of this volume gives a most valuable survey of the whole French effort in fluid mechanics, containing impressive accounts, fully illustrated, of the work of the institutes for the study of the subject at Grenoble, Lille, Marseille, Paris, Poitiers, Strasbourg and Toulouse. The second part consists of a number of original papers, of which the last, by Professor R. Timman of Amsterdam, is particularly worth noticing here. It is a very full and clear account of the whole theory of unsteady two-dimensional motions of compressible or incompressible fluid about thin areofoils oscillating in an arbitrary manner.

M. J. LIGHTHILL.

Servomechanisms and Regulating System Design, Volume II, by H. CHESTNUT and R. W. MAYER. Pp. xiii + 384. (New York: John Wiley; London: Chapman and Hall, 1955.) 68*s*.

The authors have the distinction of being the first to run to two volumes on this subject. They may expect the question to be asked of whether such wordiness has been justified. Your reviewer would allow that it has. To some extent this is because the two volumes are of quite distinct characters. The first is a reassemblage of the classical servo theory (if one may so name the works of the Brown and Campbell school) while the second is a manual of advanced applied theory and practice. The latter volume is in this respect more original than the previous one and to those directly concerned with the design of control systems much more useful. The authors and their colleagues have a wide experience of designing control systems as is evidenced by the designs given in detail which are based on well-tried and sound, perhaps rather conservative, engineering methods. There are no flighty paper-designs here; the highly refined statistical methods of design are outlined but not indulged in. The authors have the wisdom to stick to the guarantee of their practical experience —surely a wise policy in this field that has sometimes suffered over-dramatization at the hands of well-meaning disciples—and are only occasionally tempted away from it by the desire for completeness. They have successfully included much practical advice of real value to the practising expert.

The first chapter sets the practical tone of this volume by discussing measurement techniques and sine-wave generation at low frequencies for control system testing. Simple transient response analysis is suggested and an extensive treatment of approximation transformations from step-response to frequency response is given. The second chapter deals with input functions at large, exemplified by the azimuth, elevation and range functions of straight-line flight in non-dimensionalized geometric form. Rather oddly these are not transformed into frequency spectra but are dealt with in terms of equivalent sine waves. Errors to arbitrary input are assessed by means of error series which can be a hazardous procedure unless the convergence properties of the series are really understood. Noise and statistical considerations are acknowledged and represented by a generalized damped-oscillatory exponential auto-correlation function: a useful simplification in many practical problems.

The third chapter fulfils a long outstanding need in the form of a well-documented treatment of mechanical considerations of the loaded servo motor. Torque, speed and power are discussed in all aspects that concern duty cycles, gear trains and mechanical time constants. Chapter four consists of a catalogue of stabilizing networks in terms of generalized parameters, salted with a few network theorems. Chapter five gives a few of the hard facts of life in respect of the d.c. amplifier, knowingly concluded with a few words on modulators, and followed by a chapter on a.c. servos for the record.

The authors then come to the subject on which they are clearly most at home and the remaining three chapters are devoted to various aspects of the non-linear system. The first of these, on perturbation methods, deals with linearization over small departures of motor characteristics, gives an insight into the complication that arises when several inter-related variables must be simultaneously controlled by a discussion of jet engine control and concludes with mention of time-variable parameters. The next chapter stands out as the best in this volume and, after restating the basis of the describing function method, illustrates this by two extensive studies : the first on saturation effects on position control systems and the second on backlash in such systems with load resonance, both important and ubiquitous problems. The transient solutions given were obtained using an analogue computer (the method of calculating transient response in terms of the describing function approximation apparently not having penetrated across the Atlantic yet) and the reader has to accept these solutions without knowledge of their accuracy. This chapter concludes with a brief mention of relay type servomechanisms and phase plane diagrams. The final chapter indicates how non-linear elements might usefully be embodied in designs where varying parameters are rife.

Here then is a volume of great practical merit in which in respect of non-linear systems some real problems have been tackled manfully. The authors have something new to say and the manner in which they have said it will be appreciated by those who are seriously concerned with control systems.

J. H. WESTCOTT.

Boltzmann's Distribution Law, by E. A. GUGGENHEIM. Pp. 61. (Amsterdam : North-Holland Publishing Co., 1955.) 5*s*. 6*d*.

Although this booklet is written for " students of physics and chemistry in their first term at a university, provided only they have had an adequate training in elementary calculus ", it will probably be the older students and their teachers who will benefit most from it.

Lucidly written, it derives properties of such systems as ideal diatomic gases, simple crystals, or a gas of electric dipoles, and discusses some of the assumptions used in the derivations.

It seems to me that the level is still too high for first-year students, but for students in their final honours year it should provide stimulating supplementary reading. The price is so low that it can become part of every student's library. One minor criticism is that figure 3 on p. 45 seems to me to need a caption.

D. TER HAAR.

CONTENTS OF SECTION A

PAGE

Dr. A. Dalgarno and Dr. J. T. Lewis. Approximate Molecular Orbitals—IV : The $3p\sigma_u$ and $4f\sigma_u$ states of H_2^+ 285

Dr. S. N. Ghoshal and Mr. A. N. Saxena. On Neutron–Proton Pairing Interaction in Heavy Nuclei 293

Dr. A. C. Hurley. On the Method of Atoms in Molecules—III : The Ground State of Hydrogen Fluoride 301

Miss H. B. Burrows, Dr. T. S. Green, Mr. S. Hinds and Dr. R. Middleton. An Investigation of Deuteron Induced Reactions by Magnetic Analysis—III : Results for Neon and Argon 310

Mr. H. S. Perlman. The Relativistic Theory of the Specific Primary Ionization and Stopping Power and its Application in Helium 318

Mr. G. N. Harding. A Measurement of the Average Number of Neutrons per Fission from Uranium and Bismuth when bombarded with 147 mev Protons 330

Dr. L. L. Green, Dr. J. J. Singh and Dr. J. C. Willmott. The Energy Levels of ^{26}Al 335

Research Notes :

Prof. S. Devons and Mr. J. H. Towle. Range Velocity Relationship for ^{7}Li Ions in Solids 345

Mr. S. Hinds and Dr. R. Middleton. Angular Distribution of Some Inelastically Scattered Deuteron Groups from ^{27}Al 347

Dr. H. R. Allan, Mr. M. Govindjee and Mr. N. Sarma. Excitation Curve for the Reaction $^{10}B(p, \alpha)^{7}Be$ 350

Letters to the Editor :

Mrs. L. P. Chalklin and Dr. J. H. Fremlin. The Ionization of Fast Neutral Deuterium Atoms 352

Dr. J. M. Baker, Dr. B. Bleaney, Mr. P. M. Llewellyn and Dr. P. F. D. Shaw. Nuclear Spins and Magnetic Moments of Cobalt 56 and 57 . 353

Mr. S. R. Broadbent. An Examination of Observed Regularities in Nuclear Energy Levels 354

Dr. A. Lagerqvist, Mr. N. E. Lennart Nilsson and Dr. R. F. Barrow. On a Supposed Predissociation in the Spectrum of AlO 356

Reviews of Books 358

Contents of Section B 364

Figure 1. Reflection pattern from etched (110) face of copper; [$\bar{1}$10] azimuth.

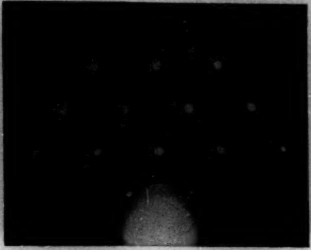


Figure 3. Pattern from copper after passing current at 20 μA cm^{-2} for 1 hour in cobalt bath; [$\bar{1}$10] azimuth.

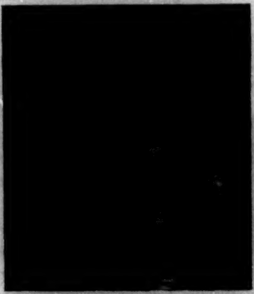


Figure 4. Pattern from etched copper covered by a cobalt layer 100 Å in thickness deposited at 160 μA cm^{-2}; copper [$\bar{1}$11] azimuth. Arced spots due to hexagonal cobalt [001] azimuth.

Figure 5. Pattern from etched copper covered by a nickel layer 100 Å in thickness deposited at 600 μA cm^{-2}; [$\bar{1}$10] azimuth.

Figure 1. Pencil line on polished brass surface. Lateral magnification 1200; viewing angle 4 degrees, specimen rotated by 0·5 degree; apparent viewing angle 14 degrees.

Figure 2. Scratch on duralumin surface made by slanting razor blade. Lateral magnification 1200; viewing angle 4 degrees, large rotation of specimen (about 2 degrees).

Magnetization Curves and Domain Structure

By L. F. BATES and A. HART†
Department of Physics, University of Nottingham

MS. received 6th *October* 1955, *and in revised form* 30th *November* 1955

Abstract. For the two main faces of a Néel-cut crystal of silicon iron a detailed study is made of the changes in domain patterns occurring when the crystal dimensions are changed. The contribution of closure domain systems to bulk magnetization along a [110] direction is treated experimentally and theoretically; the results explain differences between the ideal (Akulov) magnetization curve and curves obtained experimentally, and prove that a substantial hysteresis loss occurs in high fields with crystals of finite size.

---

§ 1. Introduction

One of the aims of modern domain theory is to account for the technical magnetization curves of ferromagnetic materials. The problem is complex, and attention has therefore been confined primarily to single crystal specimens of special crystallographic orientations and to single domain particles. Calculations by Akulov (1929, 1931) predicted the shape of single crystal magnetization curves in directions of symmetry, and these were followed by the work of Lawton and Stewart (1948) for arbitrary directions of field with respect to the crystallographic axes. The work of Lawton and Stewart was paralleled by that of Néel (1944) who went further than the previous workers in that he proposed a definite domain structure for the specimen under examination and predicted how this structure changed with increasing bulk magnetization. The domain structure considered by Néel is shown in figures 1(*a*) and (*b*), for a specimen known as the Néel-cut.

In fields below about 40 Oe, there are considerable discrepancies between the calculated, theoretical curves and experiment. Thus the inherent assumption that wall movements are reversible requires infinite initial susceptibility and zero hysteresis loss. There are also sharp kinks in the theoretical curves, which are not observed experimentally. Some of these difficulties were recently resolved by Lee (1953) who re-calculated the theoretical magnetization curve for the specific domain structure, including the small scale closure domains found on the (110) faces, of the Néel-cut specimen. He obtained an (I, H) curve which, in low fields, was depressed considerably below that of Lawton and Stewart, to an extent depending directly on the volume of the closure domains. Lee's curve agrees more closely with certain experimental measurements (cf. Williams 1937) than does that of Lawton and Stewart.

In his calculations, Lee used the theoretical domain sizes derived from Néel's original analysis. One result of this approach is to predict that the depression of the new theoretical curve below that of Akulov or of Lawton and Stewart depends on the specimen depth, L of figure 1(*a*), perpendicular to the

---

† Now at Post Office Research Station, Dollis Hill, London N.W.2

(110) faces. The depression should increase in inverse proportion to the square root of the depth, since Néel's analysis shows that as L decreases the volumes of the closure domains increase relative to the main domain volumes. No variation of the shape of the magnetization curve with specimen depth had been recorded experimentally at the time of Lee's publication.

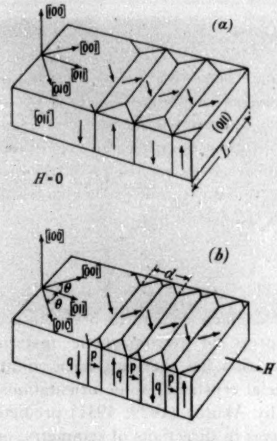


Figure 1. Domain structure of Néel-cut specimen.

Bitter figure experiments (Bates and Neale 1950, Bates and Mee 1952, Williams, Bozorth and Shockley 1949) showed that Néel's ideas are essentially correct. The patterns on the (100) faces of Néel-cut specimens indicated a domain structure exactly similar to the diagram of figure 1(*b*), while the (110)-plane patterns differed somewhat from those predicted. The q domains of figure 1(*b*) appear distinctly in the powder patterns, but in place of the p domains a complicated 'free pole' structure is observed (Bates and Mee 1952). The measured domain size varies with the magnetic field much as Néel suggested, but the measured domain width (d of figure 1(*b*)) is always greater than the theoretical value, sometimes by as much as a factor of three. As the volumes of the closure domains depend directly on the main domain widths, the theoretical volumes used by Lee in his calculations are smaller than those observed, and so Lee's corrections to the Akulov curve are correspondingly too small.

During measurements of the variation of domain sizes with magnetic field, it was noted that alterations in domain size following field changes are not always reversible. For example, let a specimen be put into an equilibrium state by reversal of a high field and its domain size be measured. If the field is now reduced without reversal, the domain width remains at the high field equilibrium value even when the applied field is substantially reduced. Thus the magnetization changes although the domain walls remain fixed. Consequently, for a chosen effective field the volumes of the closure domains for the ascending and descending branches of a magnetization loop are different, and accordingly hysteresis should be indicated in the magnetization curves calculated from experimental domain sizes.

In view of the above, the following experiments were made with a Néel-cut specimen. The Bitter figure technique was used to determine (*a*) the variation of domain size with effective magnetic field along a [110] direction, and (*b*) the effect of the specimen depth *L* on this variation. The experimental results were used to calculate theoretical magnetization curves, and to obtain an estimate of the hysteresis loss due to the effect noted above. The curves were compared with experimental ones, and with those based on theoretical domain sizes as in Lee's calculations.

§ 2. EXPERIMENTAL ARRANGEMENTS

A single crystal specimen of 3·8% silicon iron was prepared with the (110) plane 1·25 cm by 0·58 cm, the depth *L* perpendicular to this plane being 0·11 cm. The long axis of the specimen was a [110] direction. This specimen was mechanically and electrolytically polished and powder patterns could be observed on both the (100) and (110) surfaces. Measurements were mainly restricted to the patterns on the (110) face, observation of the (100) plane patterns serving merely to confirm that there were no fundamental changes in basic domain structure.

The specimen was placed between electromagnet pole-pieces of rectangular cross section, with its ends in light contact with the pole faces. The periodicity *d* of the patterns on the (110) face was observed at several points on the initial magnetization curve. At each point, the specimen was put into a cyclic state by reversing the field some twenty times, and the patterns were photographed and measured. After reaching the highest field value used, the latter was reduced without reversal, and photographs of the patterns at several field strengths down to zero were obtained. The field strengths were measured with a magnetic potentiometer used in conjunction with a galvanometer amplifier for adequate sensitivity. Measurements were made at specimen depths in the range 0·11 cm to 0·065 cm. At depths less than 0·06 cm a modified domain structure was observed, which precluded calculations of the type required in the theoretical analysis.

§ 3. GENERAL FEATURES OF THE PATTERNS

As the crystal depth *L* was reduced from 0·11 cm to 0·065 cm, the types of pattern on the (110) face remained unaltered. The photographs of figures 2(*a*) to 2(*f*) (Plate) are a typical series. In figures 2(*a*) to 2(*c*) the applied field was increasing; the specimen was later saturated, and figures 2(*d*) to 2(*f*) show the changes in pattern when the field was reduced from the saturation value to a low one. In figures 2(*a*) and 2(*b*) the clear regions correspond to the q domains, while the darker regions are 'free pole' patterns, occurring in the positions of Néel's p domains. In figure 2(*c*) the darker regions represent the backbones of the p domains which actually spread over nearly all the surface, including the relatively lighter regions. These patterns are exactly similar to those of Bates and Mee (1952), recorded in an earlier investigation of this form of specimen in the range $L=0{\cdot}6$ cm to $L=0{\cdot}2$ cm. The patterns show that Néel's p domains are replaced by a more complex structure whose exact nature is still uncertain. However, it must give a magnetization more nearly parallel to the applied field than the main domain magnetization, and in the calculations given later Néel's simple structure is used. As the chief modifications to the calculated curves arise from the contributions of the q domains in low fields, this simplification is permissible for our present purpose. The patterns on the (100) face were similar to those reported by Bates and Neale (1950), and their spacings corresponded to those required by the (110)-plane pattern.

The patterns on the (110) surface in low fields when the specimen was 0·06 cm deep were peculiar, for although they were regularly spaced, the lines of the pattern did not lie accurately along a [100] direction, but at an angle of about 5° thereto. A check on the specimen's orientation at this stage showed that its main faces were (110) planes to within 0·25°, and that its edge was certainly less than 1° from a [011] direction. Further observations as the specimen depth was reduced showed lines deviating even further from the [100] direction, until for a depth of 0·035 cm the greater part of the surface was covered with lines parallel to the edge of the specimen, i.e. to the field direction. Further information about the behaviour of such patterns will be submitted for publication in due course. They are mentioned here in order to explain why calculations made from experimental measurements of line spacing are not given for specimens less than 0·065 cm deep.

§ 4. CURVES OF DOMAIN SPACING

The measured line spacings *d* as the field was varied, with specimens of depth 0·113, 0·103, 0·093, 0·072 and 0·065 cm, were plotted against the corresponding effective fields *H*. The curve for a specimen of depth 0·072 cm is reproduced in figure 3, which is typical of all the stated depths. The broken line represents the theoretical curve calculated from Néel's formula for the chosen depth. The general characteristics of this theoretical curve, namely an initial sharp dependence of *d* on *H* in low fields, followed by a region in which the variation of *d* with *H* is very slight, are reproduced by the experimental curves. However, in no case is there a minimum in the experimental (*d*, *H*) curves. The measured spacing decreases monotonically as *H* increases and approaches an 'asymptotic' value in high fields. Moreover, the observed values of *d* are greater than the theoretical ones over the entire field range.

To this general statement regarding the numerical values of the measured and theoretical line spacings, two notes may be added: (*a*) for a given specimen depth the ratio of observed to theoretical spacing increases as the field increases;

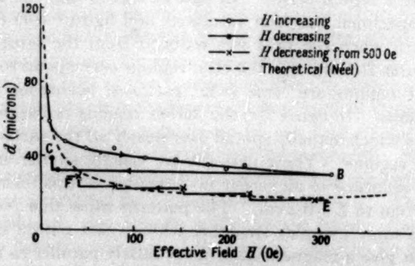


Figure 3. (*d*, *H*) curve for specimen of depth 0·072 cm.

(*b*) for a given field value this ratio decreases as the specimen depth decreases. Thus, for a depth 0·11 cm the ratio increases from 1·6 in fields of about 100 Oe to 2·0 in fields of about 300 Oe; for a specimen depth 0·07 cm the ratio increases from 1·2 to 1·8 over the same field range.

The tendency for the domain size to remain at the value corresponding to the highest field, even after the field has been considerably reduced, is shown clearly in figure 2 and also in the graph of figure 3. Because of this effect the domain size at a chosen field value is always smaller when the field is decreasing then when it is increasing. The difference between the two 'equilibrium' domain sizes increases as lower fields are approached. There is therefore a marked dependence of domain size on the maximum field employed during a cycle of magnetization. For example, when in figure 3 the line spacing at a field of about 300 Oe was measured and the field was then reduced, the domain size remained constant until the field was about 20 Oe, giving the branch BC of the graph. On the other hand if, after taking the measurement at 300 Oe, the field was increased until the specimen was saturated and then reduced until line deposits were again observed, the resulting graph, EF of figure 3, was lower than before. There is thus a wide variation in the (d, H) curve for H decreasing, unless a fixed maximum field is employed. In all the magnetization cycles used for the calculations given later, a maximum field of 300 Oe was used, so that the specimen was never saturated and field reduction was always commenced while the line structure was still visible.

§ 5. THEORETICAL ASPECTS

When a field H is applied along the [110] direction the Akulov approach gives a magnetization $I = I_s \cos\theta$ where

$$HI_s/4K = \tfrac{1}{2}\cos\theta(2\cos^2\theta - 1). \qquad \ldots\ldots(1)$$

Assuming the Néel structure of figure 1(*b*) and following Lee's (1953) treatment, the Akulov equation may be corrected to take account of the p and q domains. The correction ΔI is given by

$$\Delta I = [I_s V_p - I_s(V_p + V_q)\cos\theta]/\tfrac{1}{2}Ld \qquad \ldots\ldots(2)$$

where L is the depth of the crystal, d the spacing of a main domain pair and V_p and V_q are respectively the volumes of the p and q domains. Now, we know that in fact the p domains have a more complicated structure than that given in figure 1(*b*), but its effect is considered to be small in low fields, where V_p is small and ΔI is large and negative. In high fields ΔI is positive but small. Theoretically, for a field of about 170 Oe, $V_p = (V_p + V_q)\cos\theta$ and $\Delta I = 0$.

In the theoretical equilibrium state pictured by Néel the volumes of the closure domains are

$$V_p = \frac{\gamma L}{4W_p}\,\frac{W_q}{W_p + W_q}\,\frac{\sin\theta}{1 - \cos\theta}, \qquad \ldots\ldots(3)$$

$$V_q = \frac{\gamma L}{4W_q}\,\frac{W_p}{W_p + W_q}\tan\theta, \qquad \ldots\ldots(4)$$

while

$$d^2 = \frac{\gamma L(W_p + W_q)}{W_p W_q} \qquad \ldots\ldots(5)$$

where W_p and W_q are respectively the energy densities of the p and q domains. If V is the volume of a primary domain then

$$V + V_p + V_q = \tfrac{1}{2}Ld. \qquad \ldots\ldots(6)$$

The last five equations give $\Delta I \propto L^{-1/2}$. Now, the spacing observed in a field H may differ appreciably from that predicted by Néel's theory, when so too will V_p and V_q. Since V_p and V_q are proportional to d^2, it follows that

$$\Delta I \propto d/L. \qquad \ldots\ldots(7)$$

It seems reasonable to assume that, in a strain-free specimen, θ is given by the theoretical value of equation (1). Assuming further that for a given field H the ratio V_p/V_q is approximately the same in the actual and in the theoretical cases, from equation (2) it follows that the constant of proportionality in equation (7) is the same in both cases. Hence, if for a crystal depth L the observed pair spacing is d' and the corresponding theoretical spacing is d, then in an obvious notation

$$\Delta I'/\Delta I = d'/d. \qquad \ldots\ldots\ldots(8)$$

When therefore, due to hysteresis, different spacings d' and d'' are observed for the same field, the difference between the intensities of magnetization is

$$\Delta I' - \Delta I'' = \delta I' = \Delta I(d'-d'')/d \qquad \ldots\ldots(9)$$

and for two specimens of depths L_1 and L_2 we have from (7)

$$\frac{\delta I_1'}{\delta I_2'} = \frac{d_1'-d_1''}{d_2'-d_2''}\,\frac{L_2}{L_1}. \qquad \ldots\ldots(10)$$

The last four equations show the modifications to be made to the theoretical calculations to allow for the differences between the experimental and the theoretical domain sizes.

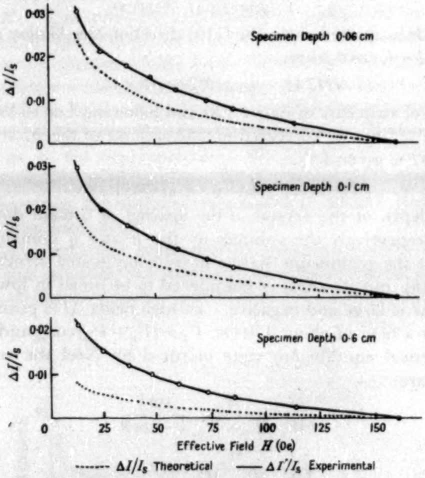


Figure 4. Comparison of $\Delta I/I_s$ and $\Delta I'/I_s$ values for three specimen depths.

Let us now consider the results of the application of this analysis to the information given by the (d, H) curves. The dotted curves of figure 4 were obtained from calculations based on Néel's theoretical domain sizes. They are for the three separate specimen depths, 0·6 cm, 0·1 cm and 0·06 cm. The full line graphs were calculated from the observed domain sizes on specimens with the corresponding values of L. The results for the specimen 0·6 cm deep are taken from a curve by Mee (1952) and usefully extend the range of the calculations. Comparing curves of each pair, the $\Delta I'$ values are appreciably greater than the ΔI values but, in going from one pair to the next, the variation in $\Delta I'$ is much less than that in ΔI. This is due to the effect of the d'/d factor in equation (8).

§ 6. QUANTITATIVE ASPECTS

The above calculations give values of $\Delta I'$ which, at a specimen depth of 0·1 cm, are twice as large as the ΔI values deduced from the theoretical domain sizes. Moreover, at any field strength the value of $\Delta I'$ does not vary as sharply with specimen depth as does the theoretical ΔI.

Lee (1953) compared his results for ΔI with the experimental magnetization curves obtained by Williams (1937) for a specimen of depth $L = 0{\cdot}23$ cm. Good agreement was obtained, but the theoretical curve was still above the experimental one. The present work indicates that the actual domain size for this specimen depth is slightly greater than twice the theoretical size, so that the 'theoretical' curve is considerably lowered, and thus gives better agreement with Williams' measurements.

A magnetization curve for a single crystal specimen with dimensions 2·1 cm by 0·9 cm in the (110) plane and a depth of 0·072 cm was obtained. Powder patterns on this specimen agreed very closely with those on the specimen used for the rest of the investigation, but the greater surface area allowed more accurate measurement of the field by the magnetic potentiometer. The search coil for measuring the corresponding induction was a single turn of No. 44 s.w.g. enamelled copper wire; both were used in conjunction with a galvanometer amplifier. This curve was first reduced to an I/I_s curve, and then experimental values of ΔI were obtained by subtraction of this I/I_s curve from the Akulov curve. The top curve of figure 5 represents the values of ΔI plotted against the field H. For comparison, the values of ΔI calculated from the theoretical domain sizes are shown by the dotted curve, and the $\Delta I'$ values obtained from the observed domain sizes are also shown. The latter curve can be extended below 10 Oe by using theoretical domain sizes since, in this low field region, the ratio d'/d decreases and approaches unity at all specimen depths. The improved agreement obtained by using the $\Delta I'$ values is now apparent.

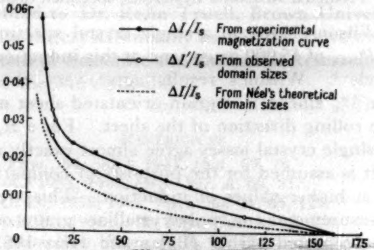


Figure 5. Comparison of $\Delta I/I_s$ and $\Delta I'/I_s$ values for specimen depth 0·072 cm.

When both the (d, H) and (d'', H) curves are used for calculating values of $\Delta I'$ and $\Delta I''$, curves such as those of figure 6 are obtained, and from them hysteresis losses can be calculated. The hysteresis loss in the range 10 to 200 Oe was calculated for various specimen depths by evaluating the appropriate areas from curves of δI against H, obtained by the substitution in equation (9) of the values of d'/d and d''/d obtained from the (d', H) and (d'', H) curves. Above 200 Oe the loss so calculated is very small, while below 10 Oe the values of d'/d deduced from the latter curves are liable to error because of the rapid variation of the spacings with H. The curves of ΔI against H showed that the hysteresis increased with decrease in specimen depth, but the changes in the values of d'/d and d''/d as the depth decreased made the variation of the hysteresis loss with depth much less than would otherwise have been expected. The loss predicted was about 1200 ergs cm^{-3} cycle^{-1} over the field range studied.

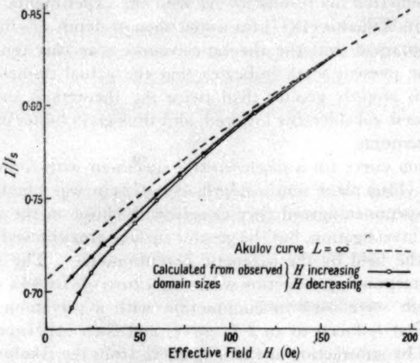


Figure 6. Predicted structural hysteresis, specimen depth 0·1 cm.

The results of Wilson (1946) on a single crystal specimen of 2·1% silicon iron extend up to a $B_{\max}$ of 15 000 gauss, and at this induction the measured loss is 3100 ergs cm^{-3} cycle^{-1}. Wilson's results agree very closely with those of Brailsford (1948) for 3% silicon iron grain-orientated sheet material magnetized at right angles to the rolling direction of the sheet. For a $B_{\max}$ of 14 000 gauss the polycrystal and single crystal losses agree almost exactly and are about 3000 ergs cm^{-3} cycle^{-1}. It is assumed for the purposes of comparison that the losses are still comparable at higher values of induction. This view is supported by powder pattern measurements on polycrystalline grain orientated material, some of which were reported earlier (Bates and Hart 1953). At a $B_{\max}$ of 18 000 gauss, the measured loss is about 4500 ergs cm^{-3} cycle^{-1} for 3·5% silicon iron magnetized in a [110] direction. The losses measured by both Brailsford and Wilson are twice as great in the [110] direction as in the [100] direction, while the losses in the [111] and [110] directions are approximately equal over the entire range of measurement. It has been observed that the losses in grain-orientated sheet material increase as the sheet thickness decreases and as the grain size is reduced (cf. Yensen and Ziegler 1935, 1936).

§ 7. DISCUSSION OF RESULTS

It is now well established that the overall domain structure of any specimen is largely influenced by the closure domains formed in its surface layers and that these have much bearing on susceptibility values and hysteresis phenomena.

In this paper we have dealt with an effect of closure domains on the mean magnetization of a Néel-cut specimen. This problem was first studied by Lee who based his calculations on the purely theoretical formulation of the domain structure by Néel. In this paper we have extended Lee's results by an experimental study of the relevant domain structures. Lee found that the theoretical (I, H) curves given by his calculations were more in agreement with experimental (I, H) curves than were those calculated by Akulov. It has here been shown that by basing the calculations on observed domain structures, instead of on those postulated by Néel, still better agreement with experiment is obtained. The improvement is mainly confined to field ranges up to about 100 Oe, and are particularly noticeable in low fields, although we know that the full implications of the closure domain structure on the (110) plane have not yet been explored.

The hysteresis loss in the upper 'rotation' region of the magnetization curve, predicted in § 5, must be due basically to wall impedance, and as it depends so markedly on closure domain structure it might be termed a closure domain hysteresis. Wall impedance hysteresis models have hitherto been confined to low field regions of the order of the coercivity, where of course wall impedance is extremely important. The point emphasized here is that the calculated hysteresis in the 'rotation' region of the magnetization curve due to this previously ignored effect corresponds to a quarter of the d.c. losses measured in the [110] direction. As Lee pointed out, the effects of closure domains are such that so-called 'constants' deduced from technical magnetization curves of ferromagnetic materials may be determined by specimen shape as well as by the nature of the material itself. We now see that this statement also applies to energy changes during the magnetization of single crystals.

ACKNOWLEDGMENTS

We are much indebted to Dr. Rohn Truell, Brown University, Providence, U.S.A., who kindly gave the crystals mainly used in the above work. A. H. wishes to thank the Department of Scientific and Industrial Research for a maintenance allowance which made the work possible.

REFERENCES

AKULOV, N. S., 1929, *Z. Phys.*, **54**, 582 ; 1931, *Ibid.*, **69**, 78.
BATES, L. F., and HART, A., 1953, *Proc. Phys. Soc.* A, **66**, 813.
BATES, L. F., and MEE, C. D., 1952, *Proc. Phys. Soc.* A, **65**, 129.
BATES, L. F., and NEALE, F. E., 1950, *Proc. Phys. Soc.* A, **63**, 374.
BRAILSFORD, F., 1948, *J. Instn Elect. Engrs*, Pt. I, **95**, 522.
ELSCHNER, B., 1953, *Ann. Phys., Lpz.*, **13**, 290.
LAWTON, H., and STEWART, K. H., 1948, *Proc. Roy. Soc.* A, **193**, 72.
LEE, E. W., 1953, *Proc. Phys. Soc.* A, **66**, 623.
MEE, C. D., 1952, *Ph.D. Thesis*, University of Nottingham.
NÉEL, L., 1944, *J. Phys. Radium*, **5**, 241, 265.
WILLIAMS, H. J., 1937, *Phys. Rev.*, **52**, 747.
WILLIAMS, H. J., BOZORTH, R. M., and SHOCKLEY, W., 1949, *Phys. Rev.*, **75**, 155.
WILSON, A. J. C., 1946, *Proc. Phys. Soc.*, **58**, 21.
YENSEN, T. D., and ZIEGLER, N. A., 1935, *Trans. Amer. Soc. Metals*, **23**, 556 ; 1936, *Ibid.*, **24**, 337.

The Absorption Edge of Amorphous Selenium and its change with Temperature

By C. HILSUM

Services Electronics Research Laboratory, Baldock, Herts.

MS. received 7th November 1955, and in final form 21st December 1955

Abstract. Measurements have been made of the absorption coefficient of amorphous selenium in the region from 0·58 to 0·66 microns. The effect of temperature on the absorption edge has also been determined, and the edge is shown to shift by 2·7 Å deg^{-1}, equivalent to a change in the energy gap of $9{\cdot}7\times10^{-4}$ ev deg^{-1}.

---

§ 1. Introduction

In recent years amorphous selenium has been found useful as a photoconductor (Weimer and Cope 1951) and an infra-red optical material (Gebbie 1952) and several measurements have been made of the refractive index and absorption coefficient both in the visible and in the infra-red. The various determinations of refractive index (Dowd 1951, Saker 1952, Wood 1902) have given good agreement, and we may consider this quantity as well defined, but wide divergence is shown between the values obtained by different workers for the absorption coefficient, particularly in the region of the absorption edge, 0·50 to 0·70 microns, where the coefficient varies very rapidly with wavelength. Criticism was directed at some of the earlier experiments (Wood 1902, Becker and Schaper 1944) on the grounds that the specimens used were obtained by sputtering and probably contained some crystalline material; for this reason the two most recent determinations, those of Dowd (1951) and Gilleo (1951), using evaporated films, were to be preferred. Though the agreement here is considerably better, there is still a discrepancy between the two sets of results, Dowd's values being almost 30% lower than Gilleo's. A new determination has now been made using different experimental methods.

A further point which warrants investigation is the variation of the absorption coefficient with temperature, an effect which causes the absorption edge to shift to longer wavelengths as the temperature increases. The only detailed measurements published (Gilleo 1951, Saker 1952), were made at wavelengths near and beyond 0·65 microns, where the absorption coefficient is not high at room temperature; these have now been extended to shorter wavelengths.

§ 2. Preparation of Specimens

The layers used were made by evaporating 'specpure' selenium shot, obtained from Johnson Matthey Ltd., from an electrically heated tantalum boat in a vacuum of less than 5×10^{-5} mm Hg. The substrates, polished glass discs 2 in. in diameter, were placed approximately 25 cm from the evaporator to obtain films of uniform thickness, but even so it was found that rapid evaporation resulted in a non-uniform deposit. The temperature of the boat was therefore adjusted to give a deposition rate of about 100 Å per minute. At each evaporation four specimens were prepared, one in the form of a thickness monitor from which the approximate thickness of all four layers could be determined, using Tolansky multiple beam fringes. The three other specimens were examined for blemishes and pinholes, and the best chosen for transmission measurements. Fourteen layers were used, ranging in thickness from 0·7 to 3 microns.

§ 3. DETERMINATION OF ABSORPTION COEFFICIENT

3.1. *Transmission Measurements*

The transmission of each specimen was measured over the wavelength range 0·5 to 1·0 microns, using a General Electric recording spectrophotometer. It is estimated that the transmissions are accurate to 0·5% and the wavelengths to within 10 Å. The temperature of the layer was noted during the run, and for all specimens it was within 2 degrees of 300°K.

A typical curve is shown in figure 1. It can be seen that the curve divides into two parts—a short wavelength region where the transmission increases rapidly with wavelength, and a second region where the transmission oscillates with wavelength, due to interference effects in the layer. From the second region an accurate measure of the thickness of the specimen is made in the following way. The maxima occur at wavelengths where the thickness d is related to the wavelength by the relation

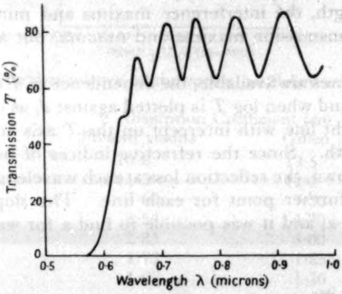


Figure 1. A typical transmission–wavelength curve: Film thickness = 1·57 microns.

$$2nd = N\lambda \qquad \ldots\ldots(1)$$

where n is the refractive index and N an integer. Similarly, at the minima

$$2nd = (N + \tfrac{1}{2})\lambda. \qquad \ldots\ldots(2)$$

From the published data on refractive index (Saker 1952), values of $2n/\lambda$ are tabulated. From equations (1) and (2), it is apparent that the maxima and minima are spaced at equal intervals of $2n/\lambda$, and from the mean value of this interval d can be calculated.

An alternative procedure involves a prior knowledge of the approximate value of d. The thickness monitors mentioned in the previous section give this for each specimen, and the oscillatory pattern to be expected from such a thickness is determined, using the tables of $2n/\lambda$, and compared with that actually occurring. The deviations between the two curves indicate how the actual thickness differs from the approximate value, and by making a succession of small changes in d it is possible to find accurately the thickness necessary to generate the pattern observed. This method is more accurate than the other, particularly for the thinner films, where only a small number of maxima are found in the wavelength range recorded, but in practice the two methods gave answers agreeing to within about 2%.

It is believed that the interference method of measuring thickness is more accurate than that used by Dowd and Gilleo, who weighed the selenium in their layers and deduced thickness from the published density of amorphous selenium (4·26 to 4·30 gm cm^{-3}).

To calculate the absorption coefficient at any wavelength λ, the variation of transmission T with layer thickness must be known, and this can normally be derived from the T–λ recordings. At some wavelengths, however, interference effects complicate the interpretation, and a true transmission curve must be found. This is done by drawing the envelope of points of maximum positive departure from the smooth curve, and the corresponding envelope for points of maximum negative departure. The mean of the two envelopes is the true transmission curve. It should be noted that on a curve where transmission varies rapidly with wavelength, the interference maxima and minima do not coincide with the observed transmission maxima and minima, but are slightly displaced from them.

Once true T–λ curves are available, the dependence of T on d can be tabulated for each wavelength and when log T is plotted against d, as in figure 2, the points should lie on a straight line with intercept on the T axis equal to the reflection loss at that wavelength. Since the refractive indices of both the film and the glass substrate are known, the reflection loss at each wavelength can be computed, and this provides a further point for each line. The slope of the line is the absorption coefficient α, and it was possible to find α for wavelengths from 0·59 to 0·66 micron.

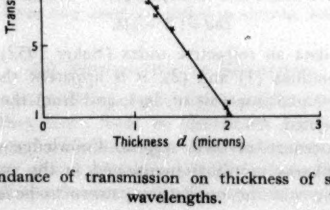


Figure 2. The dependance of transmission on thickness of specimen at two typical wavelengths.

As a check, the variation of transmission with thickness at 0·589 microns was found by a separate experiment, using a sodium lamp and a quartz double monochromator, and the same selenium specimens. The results were in good agreement with those obtained using the recording spectrophotometer.

3.2. *Values for Absorption Coefficient*

In the table the results obtained are compared with data kindly supplied by Dr. Gilleo and Mr. Dowd, and a plot of the results is shown in figure 3.

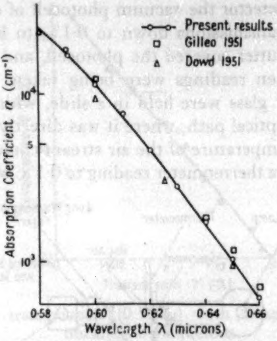


Figure 3. The dependence of absorption coefficient on wavelength.

Wavelength (microns)	Absorption Coefficient ($cm^{-1} \times 10^3$) Present results	Gilleo	Dowd
0·59	16·8	19·5	15·0
0·60	11·3	12·2	9·5
0·61	7·60	7·60	6·0
0·62	4·64	4·70	3·8
0·63	2·83	3·00	2·5
0·64	1·73	1·85	1·6
0·65	1·03	1·20	0·98

It can be seen that there is fair agreement with Gilleo over most of the wavelength range covered, but at several points the divergence is greater than our estimated experimental error of ± 5%.

§ 4. EFFECT OF TEMPERATURE ON THE ABSORPTION COEFFICIENT

4.1. *Experimental Details*

At a wavelength where the absorption coefficient is α_0, the transmission of a film of thickness d is given by

$$T = T_0 \exp(-\alpha_0 d) \qquad \ldots\ldots(3)$$

where T_0 takes account of the reflection loss. On differentiating with respect to temperature θ

$$dT/d\theta = -Td(d\alpha/d\theta) \qquad \ldots\ldots(4)$$

as T_0 varies much less with temperature than T. If we define $-(dT/d\theta)/T$ as F, and $d\alpha/d\theta$ as γ, we have

$$F = d\gamma. \qquad \ldots\ldots(5)$$

Equation (5) indicates that if F can be determined for a series of specimens of different thicknesses, a graph of F against d should be a straight line through the origin, and the slope is equal to γ.

Therefore the method adopted was to find, using monochromatic light, the T–θ curves over the temperature range 20°c–40°c for the specimens mentioned above, deduce F at the mean temperature, 30°c, for each specimen, and finally to measure γ from the F–d curve.

The apparatus used is represented in figure 4, the source being a 20 watt sodium lamp, and the detector the vacuum photocell of a Baldwin transmission photometer, enabling transmissions down to 0·1% to be measured with good accuracy. A camera shutter covered the photocell, and to avoid fatigue effects it was opened only when readings were being taken. The specimen and a comparison disc of clear glass were held in a slide, which could be adjusted to place either disc in the optical path, where it was directly in an air blast from an electric blower. The temperature of the air stream could be adjusted by means of an electric heater, and a thermometer reading to 0·1°c was placed in the stream.

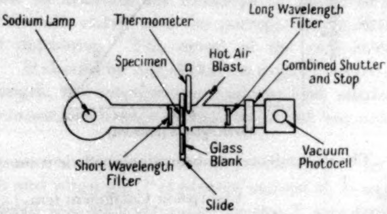


Figure 4. Apparatus for measuring temperature coefficient of transmission at 5893 Å.

To lessen the intensity of the green and red sodium lines, a short and a long wavelength filter were necessary. The green line was reduced to a negligible amount by a 1 micron selenium layer, and the red line was weakened by an N.P.L. photocell correction filter (Harding 1950). A spectrographic test showed that when both these filters were in position, the red line could be neglected for films thinner than 2 microns.

The procedure employed was to measure transmission at temperature intervals of two or three Centigrade degrees, first with temperature increasing in steps from room temperature to 40°c, and then with it being reduced to the original level. For each specimen the cycle was repeated at least once. Checks showed that the layer needed about three minutes to stabilize in temperature after the heater had been adjusted.

4.2. *Results*

It was found that for all except the thinnest specimen, the $T-\theta$ plots were good approximations to straight lines. Since the temperature range covered was only ±10°c, it is doubtful if any physical significance can be given to this point. A typical graph is shown in figure 5. The circles represent the mean of between two and four experimental observations, so that the line is identified by about 30 measurements. The transmission–temperature graph for the thinnest layer, 0·70 microns thick, showed a small departure from linearity, the slope tending to decrease in magnitude at the higher temperatures.

From the graphs, T and $dT/d\theta$ were measured at 30°c, and F plotted against d, as in figure 6; the points for layers thinner than 2 microns lie on a straight line through the origin, and the slope indicates a value for $d\alpha/d\theta$ of 190 cm^{-1} deg^{-1}. Results for thicker films lie below the straight line, showing a proportionately smaller modulation of transmission than the thinner specimens. It is probable that this effect is due to the red sodium line, which becomes of importance only for such thick films.

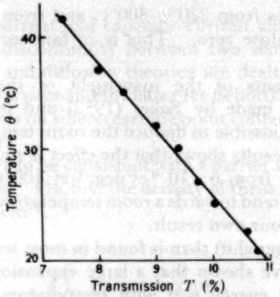


Figure 5. Variation of transmission (0·589 μ) with temperature for selenium film, thickness 0·98 microns.

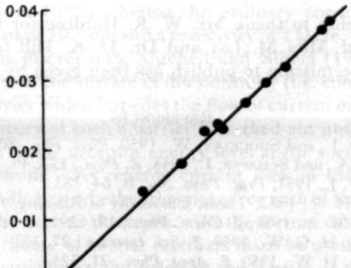


Figure 6. Dependence on film thickness of $(dT/d\theta)/T$.

The change of absorption coefficient with temperature is now known, and the change of absorption coefficient with wavelength can be read off figure 3, at a wavelength of 0·589 microns. We deduce how much the absorption edge must have shifted in wavelength to change the absorption coefficient at 0·589 microns by 190 cm^{-1}, and this shift is caused by a temperature change of one degree. It is seen that the effect of temperature is to move the edge by 2·7 Å deg^{-1}, equivalent to an energy change of $9{\cdot}7 \times 10^{-4}$ ev deg^{-1}.

§ 5. DISCUSSION

It appears from figure 3 that the absorption edge in amorphous selenium is not steep enough to deduce accurately the activation energy E, though we may conclude that the energy gap is between 1·9 and 2·1 ev. As a result the conductivity at room temperature is very small, and it would be difficult to determine E by electrical methods. However, since amorphous selenium is really a supercooled liquid, it might be expected to have properties similar to those of liquid selenium. Henkels (1950) has made electrical measurements on liquid selenium at temperatures from 220°C–500°C, and from these it can be deduced that E is 2·3 ev at absolute zero. This is in fair agreement with the optical determination.

Previous measurements of the movement of the absorption edge with temperature have been made by Saker (1952) and by Gilleo (1951). From neither experiment is it possible to deduce the room temperature shift accurately, but an analysis of their results shows that the effect is smaller at low temperatures than at high, increasing from 6×10^{-4} ev deg^{-1} at 100°K to 16×10^{-4} ev deg^{-1} at 650°K. Both sets of data tend towards a room temperature value near 10^{-3} ev deg^{-1}, in good agreement with our own result.

This represents a larger shift than is found in most semiconductors. Bardeen and Shockley (1950) have shown that a large expansion coefficient can lead to a large variation of the energy gap with temperature. In this respect it is interesting to note that the expansion coefficient of amorphous selenium, 56×10^{-6} deg^{-1} (Sato and Kaneko 1949), is one of the largest known.

ACKNOWLEDGMENTS

The author wishes to thank Mr. W. R. Harding for the preparation of the selenium films, and Miss M. Lay and Dr. D. K. Hill for assistance with the measurements. Permission to publish has been given by the Admiralty.

REFERENCES

BARDEEN, J., and SHOCKLEY, W., 1950, *Phys. Rev.*, **80**, 72.
BECKER, A., and SCHAPER, I., 1944, *Z. Phys.*, **122**, 49.
DOWD, J. J., 1951, *Proc. Phys. Soc.* B, **64**, 783.
GEBBIE, H. A., 1952, *Thesis*, University of Reading.
GILLEO, M. A., 1951, *J. Chem. Phys.*, **19**, 1291.
HARDING, H. G. W., 1950, *J. Sci. Instrum.*, **27**, 132.
HENKELS, H. W., 1950, *J. Appl. Phys.*, **21**, 725.
SAKER, E. W., 1952, *Thesis*, University of Reading.
SATO, T., and KANEKO, H., 1949, *Tech. Rep. Tôhoku Univ.*, **14**, 45.
WEIMER, P., and COPE, A., 1951, *R.C.A. Rev.*, **12**, 314.
WOOD, R. W., 1902, *Phil. Mag.*, **3**, 607.

Surface Barriers at Semiconductor Contacts

By R. STRATTON

Research Department, Metropolitan-Vickers Electrical Co. Ltd., Manchester

Communicated by B. G. Churcher; MS. received 18*th August* 1955, *and in final form* 5*th January* 1956

Abstract. The voltage–current and capacity–current characteristics for current flow across a surface of discontinuity between two semiconductors have been derived. Both the diode and diffusion theories are dealt with, the former being treated in detail for Mott barriers and the latter for Schottky barriers. The results are applied to measurements on silicon carbide point contacts and a grain boundary in germanium respectively.

The resistance is shown to be consistent with a barrier which is caused by donor and acceptor surface states. On SiC the density of these states is about 10^{13} cm^{-2} and on Ge about 5×10^{11} cm^{-2}.

---

§ 1. Introduction

Experiment has shown that a surface of discontinuity between two semiconductors gives a high resistance for current flow: high, that is, with respect to the resistance to be expected from the bulk resistivity of the materials. Thus the resistance for low applied voltages across a SiC–SiC point contact may be as high as $10^{10}\,\Omega$ whereas the ordinary constriction resistance, assuming a contact radius of 10^{-3} cm and a resistivity of $1\,\Omega$ cm, is between 10^2 and $10^3\,\Omega$. To explain this discrepancy Mitchell and Sillars (1949) suggested the presence of surface states at the surface of discontinuity (i.e. contact surface in this case) giving rise to a barrier which impedes the flow of current carriers.

In this report the theory of such a barrier is worked out under the assumption of a single donor and a single acceptor energy level at the surface both of which may trap or emit electrons. As regards current flow in the barrier itself two possibilities arise depending on whether the mean free path of the carriers is long or short with respect to the barrier thickness. In the former case the diode and in the latter the diffusion theory of barrier rectifiers must be used. It is found that the results depend essentially on the fields in the barrier at the position of the surface. These can be related to the potential heights of the barrier if the charge distribution within the barriers is known.

The characteristics are obtained by examining the conditions for electrostatic charge neutrality and for a stationary distribution of charges on the surface states. Finally the explicit relation between the current and the voltage drop across the barrier can only be obtained if the impurity charge distributions in the barrier are specified. Two special cases were dealt with, one because it seems to apply to the SiC–SiC contacts already mentioned, the other because it is relevant to a Ge grain boundary investigated by Taylor, Odell and Fan (1952). The theory is, however, more general and can be applied to any impurity distribution. In § 4, the applicability of the present calculation to commercial, non-ohmic, bonded silicon carbide resistors, is discussed.

§ 2. THEORY

2.1. *Model of the Surface Barrier*

The model of the surface of an n-type semiconductor in equilibrium is represented by the electronic potential energy diagram of figure 1. The theory developed below can easily be applied to p-type semiconductors by exchanging the roles of electrons and donors for those of holes and acceptors. It will be assumed that two types of surface states exist, namely donor states D (energy ϵ_d below the conduction band) and acceptor states A (energy ϵ_a below the conduction band). Thus the D states are neutral when full and the A states when empty (of electrons). The charge on the surface gives rise to an internal barrier of height ϕ_0 ev. Now if

$$\left.\begin{aligned}&\exp\left(\frac{\epsilon_d}{\boldsymbol{k}T}\right),\ \exp\left(\frac{E_g-\epsilon_a}{\boldsymbol{k}T}\right) \ll \exp\left(\frac{E_g}{\boldsymbol{k}T}\right) \text{ and}\\ &\exp\left(\frac{\phi_0-\zeta}{\boldsymbol{k}T}\right) \ll \exp\left(\frac{E_g}{2\boldsymbol{k}T}\right)\end{aligned}\right\} \quad \ldots\ldots(1)$$

where E_g ev is the forbidden gap width and ζ ev is the Fermi energy in the bulk of the semiconductor relative to the bottom of the conduction band, then it can easily be shown that interactions of electrons in the D states with the valence band may be neglected while the A states may be taken as permanently occupied by electrons. Thus if F_0/K is the field at the top of the barrier inside the semiconductor (measured as a rate of change of potential energy, i.e. in ev cm^{-1}) where K is the dielectric constant then

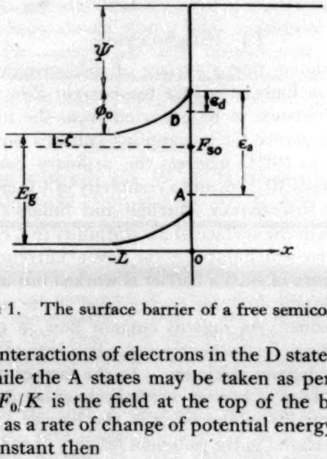


Figure 1. The surface barrier of a free semiconductor.

$$F_0/4\pi \boldsymbol{e}^2 = \tfrac{1}{2}\left[(N_a - N_d) + n_{d0}\right] \quad \ldots\ldots(2)$$

($\frac{1}{2}n_{d0}$ is the number of electrons per unit area on the D states in equilibrium, $\frac{1}{2}N_d$, $\frac{1}{2}N_a$ are the numbers per unit area of D and A states respectively). F_0 can be related to ϕ_0 and the barrier thickness L if the charge distribution in the barrier is known. Two special cases will be dealt with later.

If c is the average capture cross section of the D states and 2τ is the lifetime of electrons in them, then the condition that n_{d0} is stationary gives

$$I_0 c (N_d - n_{d0}) = \boldsymbol{e}\, n_{d0}/2\tau \quad \ldots\ldots(3)$$

where

$$I_0 = n\boldsymbol{e}\left(\frac{\boldsymbol{k}T}{2\pi \boldsymbol{m}}\right)^{1/2} \exp\left(-\frac{\phi_0}{\boldsymbol{k}T}\right) \quad \ldots\ldots(4)$$

($\boldsymbol{n}$ = electron density in the bulk of the semiconductor, T = absolute temperature) I_0 is the current of electrons approaching unit area of the surface. More exactly if $c(E_x)$ is the capture probability per D state in unit area and $E_x = p_x^2/2\boldsymbol{m}$ (p_x is the momentum in the x direction normal to the surface) then

$$n\left(\frac{\boldsymbol{k}T}{2\pi\boldsymbol{m}}\right)^{1/2} c = n\left(\frac{\boldsymbol{k}T}{2\pi\boldsymbol{m}}\right)^{1/2}\int_0^\infty \exp\left(-\frac{E_x}{\boldsymbol{k}T}\right)\frac{c(E_x)E_x}{(\boldsymbol{k}T)^2}\,dE_x. \quad\ldots\ldots(5)$$

Similarly $1/2\tau$ will be the average probability per unit time that an electron is emitted from a D state. Introducing the constants $\alpha = n_{\mathrm{d}0}/N_{\mathrm{d}}$, $\beta = (N_{\mathrm{a}} - N_{\mathrm{d}})/N_{\mathrm{d}}$ equations (2) and (3) can be rewritten as

$$\alpha + \beta = 2F_0/4\pi\boldsymbol{e}^2N_{\mathrm{d}} \quad\ldots\ldots(3a)$$

$$\alpha/(1-\alpha) = 2I_0 c\tau/\boldsymbol{e}. \quad\ldots\ldots(4a)$$

Finally, since the system is in thermal equilibrium

$$\left.\begin{aligned} \frac{n_{\mathrm{d}0}}{N_{\mathrm{d}}} &= \left[1+\exp\left(\frac{\phi_0-\zeta-\epsilon_{\mathrm{d}}}{\boldsymbol{k}T}\right)\right]^{-1} \\ \text{i.e.}\quad \alpha/(1-\alpha) &= \exp[-(\phi_0-\zeta-\epsilon_{\mathrm{d}})/\boldsymbol{k}T]. \end{aligned}\right\} \quad\ldots\ldots(6)$$

Comparing with equation (4*a*) and using the relation (valid for a non-degenerate semiconductor)

$$n = 2N_0 \exp \zeta/\boldsymbol{k}T \quad\ldots\ldots(7)$$

where $N_0 = (2\pi\boldsymbol{mk}T)^{3/2}/\boldsymbol{h}^3$ we have

$$c\tau = \frac{\exp(\epsilon_{\mathrm{d}}/\boldsymbol{k}T)}{4N_0(\boldsymbol{k}T/2\pi\boldsymbol{m})^{1/2}}. \quad\ldots\ldots(8)$$

If contact is made between two identical semiconductors of the type just described the equations remain unaltered since the quantities $\frac{1}{2}n_{\mathrm{d}0}$, $\frac{1}{2}N_{\mathrm{d}}$, $\frac{1}{2}N_{\mathrm{a}}$ and $1/2\tau$ are doubled while there are two equal terms on the left-hand sides of equations (2) and (3).

So far the mechanism of the conduction in the barrier layer itself has not been required since no resultant current flows when there is equilibrium. For an external applied potential two cases will have to be considered. If l is the mean free path of the electrons then one of the two extreme conditions $L \ll l\phi_0/\boldsymbol{k}T$ or $L \gg l\phi_0/\boldsymbol{k}T$ require the use of either the diode or the diffusion theory respectively. (The inequalities compare the average energy exchange due to a collision ($\simeq \boldsymbol{k}T$) with the gain in potential energy between two collisions ($\simeq \phi_0 l/L$).) The diode theory will be applied to the voltage–current characteristics of SiC contacts since the analysis of the experimental results (§ 3.1) indicates that $L \simeq 10^{-6}$ cm and $\phi_0 \simeq \frac{1}{2}$ ev (cf. table) while $l \simeq 5\times10^{-6}$ cm for green SiC and a factor about 10 less for black SiC as deduced from mobility measurements (Busch and Labhart 1946). Taylor, Odell and Fan have applied the diffusion theory to a grain boundary barrier in germanium but restricted the detailed analysis to the zero bias resistance. The theory will be extended to cover the high voltage region.

2.2. *General Diode Theory*

If a potential $-V/\boldsymbol{e}$ is applied across the barrier an electron current flows from left to right and the situation shown in figure 2 occurs. V_1 (ev) represents the drop in height of the barrier on the left-hand side and V_2 (ev) the rise in height of the barrier on the right-hand side, so that $V_1 + V_2 = V$.

If F_1/K, F_2/K are the fields on top of barriers 1 and 2 then (F_2 is negative)

$$(F_1 - F_2)/4\pi \mathbf{e}^2 = N_{\mathrm{a}} - N_{\mathrm{d}} + n_{\mathrm{d}} \qquad \ldots\ldots(9)$$

(n_{d} is the number of electrons in the D states).

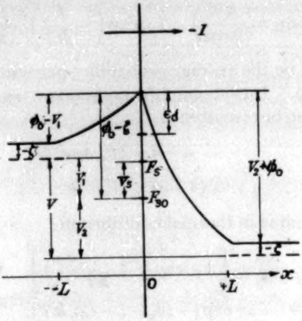


Figure 2. The barrier at the surface of discontinuity between two semiconductors.

The condition for stationary n_{d} now becomes

$$I_0 c (N_{\mathrm{d}} - n_{\mathrm{d}}) [\exp(V_1/\boldsymbol{k}T) + \exp(-V_2/\boldsymbol{k}T)] = \mathbf{e} n_{\mathrm{d}}/\tau. \qquad \ldots\ldots(10)$$

The two terms on the left-hand side of the above equation represent the electron streams from the bulk of the crystals which have sufficient energy to surmount the two barriers and are captured by the D states (diode theory).

The resultant current density from crystal 1 to crystal 2 is $-I$ (negative because of the negative charge of the electrons) where

$$I = I_0 \exp(V_1/\boldsymbol{k}T) - I_0 \exp(-V_2/\boldsymbol{k}T) = I_0 \exp(V_1/\boldsymbol{k}T)\{1 - \exp(-V/\boldsymbol{k}T)\}. \qquad \ldots\ldots(11)$$

Elimination of V_1 between equations (10) and (11) gives

$$\frac{Ic\tau}{\mathbf{e}} \Big/ \tanh \frac{V}{2\boldsymbol{k}T} = \frac{n_{\mathrm{d}}}{N_{\mathrm{d}} - n_{\mathrm{d}}}. \qquad \ldots\ldots(12)$$

By use of equation (4*a*) this can be put in the more convenient form

$$\frac{n_{\mathrm{d}}}{N_{\mathrm{d}}} = \frac{j}{\tanh(V/2\boldsymbol{k}T) + j} \qquad \ldots\ldots(13)$$

where $j = \dfrac{I}{2I_0} \dfrac{\alpha}{1-\alpha}$ and is proportional to I.

Equation (9) may be written in the more convenient form

$$\frac{F_1 - F_2}{2F_0} = \frac{\beta + n_{\mathrm{d}}/N_{\mathrm{d}}}{\alpha + \beta} \qquad \ldots\ldots(14)$$

(using equation (3*a*)). From these n_{d} may be eliminated to give

$$\frac{\alpha + \beta}{2F_0}(F_1 - F_2) = \frac{j}{j + \tanh(V/2\boldsymbol{k}T)} + \beta. \qquad \ldots\ldots(15)$$

For a complete solution of the problem the connection between F_1, F_2 and V_1, V_2 must be established. This will be discussed in § 2.4 for a particular case.

By defining a quasi Fermi level F_s for the electrons on the D states it will be shown that these electrons are very nearly in thermal equilibrium with the electrons in the interior of crystal 1. Thus

$$n_d/N_d=\left[1+\exp\left(\frac{\phi_0-\zeta-\epsilon_d-V_s}{\boldsymbol{k}T}\right)\right]^{-1}=\left[1+\left(\frac{1-\alpha}{\alpha}\right)\exp\left(-\frac{V_s}{\boldsymbol{k}T}\right)\right]^{-1} \quad \ldots\ldots(16)$$

where V_s(ev) is the amount of energy by which F_s has been raised from its original $(V=0)$ position F_{s0}.

Comparison of equations (16) and (13) shows the following relation :

$$I=2I_0\exp(V_s/\boldsymbol{k}T)\tanh(V/2\boldsymbol{k}T). \quad \ldots\ldots(17)$$

The separation of F_s from the Fermi level in the interior of crystal 1 is V_1-V_s which can be derived from equations (11) and (17). Thus

$$V_1-V_s=\boldsymbol{k}T\log\left[\frac{2}{1+\exp(-V/\boldsymbol{k}T)}\right] \quad \ldots\ldots(18)$$

which increases monotonically to $V_1-V_s\sim\boldsymbol{k}T\log 2$ as V increases from 0. For room temperature $V_1-V_s\simeq 0{\cdot}017$ ev so that the electrons on the D states are very nearly in thermal equilibrium with those in the interior of crystal 1.

2.2.1. *Modification for a non-symmetrical contact.*

The theory developed so far yields non-linear but symmetrical voltage–current characteristics according to equation (15). This is contrary to the results of all the experiments to be described later. Such a lack of symmetry is however to be expected since the various parameters of the two crystals may be different. To simplify the analysis the values of Ψ and ϵ_d respectively will be assumed equal thus giving one set of D levels only. This will not be exactly true since besides depending on the pure SiC lattice and the origin (e.g. impurity atoms) of the surface states, these parameters may depend on the crystal face at the contact. Removing this approximation would give a barrier with a discontinuous potential drop in the interface equal to $\Psi_a-\Psi_b$ and two surface states levels separated in energy by $\Psi_a-\Psi_b+\epsilon_{da}-\epsilon_{db}$ where a and b refer to the left-hand and right-hand crystals of figure 2 respectively. Both these effects will be unimportant unless the two energy differences become large.

It will however be assumed that the capture cross sections and lifetimes of the D states with respect to the two crystals are different. The justification of this point will depend on the mechanism that is responsible for the capture and emission processes. Now the lattice waves for the contact under discussion may be non-symmetrical with respect to the interface and, if these are important for the capture process, the relative direction of motion of the electron may influence the c and τ values.

The assumption of equal values for ϵ_d and Ψ leads to the two relations

$$c_a\tau_a=c_b\tau_b\equiv c\tau$$

(using equation (8)) and $\phi-\zeta_{a0}=\phi_{b0}-\zeta_b$ respectively, so that the two I_0 values are the same.

Proceeding as before the final result for the current density as a function of V is

$$(\alpha+\beta)\frac{F_a-F_b}{F_{a0}-F_{b0}}=\frac{j_a}{\mathrm{Th}\left(\frac{c_b}{c_a},\frac{V}{2\boldsymbol{k}T}\right)+j_a}+\beta \quad \ldots\ldots(19)$$

where
$$j_{\rm a}=\frac{c_{\rm a}}{c_{\rm a}+c_{\rm b}}\frac{I}{I_0}\frac{\alpha}{1-\alpha}$$

and
$$\mathrm{Th}\left(\frac{c_{\rm b}}{c_{\rm a}},\frac{V}{2\boldsymbol{k}T}\right)\equiv\frac{1-\exp\left(-V/\boldsymbol{k}T\right)}{1+(c\ /c_{\rm a})\exp\left(-V/\boldsymbol{k}T\right)}.$$

Thus the ratio j/I and the function Th depends on the direction in which V is applied.

Introducing the quasi-Fermi level $F_{\rm s}$ as before (equation (16)) we find

$$I=\frac{c_{\rm a}+c_{\rm b}}{c_{\rm a}}I_0\exp\left(V_{\rm s}/\boldsymbol{k}T\right)\mathrm{Th}\left(\frac{c_{\rm b}}{c_{\rm a}},\ \frac{V}{2\boldsymbol{k}T}\right)\qquad\ldots\ldots(20)$$

also

$$V_1-V_{\rm s}=\boldsymbol{k}T\log\left[\frac{c_{\rm a}+c_{\rm b}}{c_{\rm a}}\,\frac{1}{1+(c_{\rm b}/c_{\rm a})\exp\left(-V/\boldsymbol{k}T\right)}\right]\qquad\ldots\ldots(21)$$

$$\sim\boldsymbol{k}T\log\left\{(c_{\rm a}+c_{\rm b})/c_{\rm a}\right\}\quad\text{as } V \text{ becomes large.}$$

Thus the position of the quasi-Fermi level $F_{\rm s}$ is different for the two directions of current flow.

2.3. *General Diffusion Theory*

If U is the potential energy of an electron at the point x in the barrier then

$$I=\mu\left[n(x)\frac{\partial U(x)}{\partial x}+\boldsymbol{k}T\frac{\partial n(x)}{\partial x}\right]\qquad\ldots\ldots(22)$$

where $n(x)$ is the electron density at x and μ is the mobility. This equation must now be solved for the situation illustrated in figure 2. Applying equation (22) first to crystal 1 and then to 2

$$\left.\begin{aligned}I&=F_1\mu\left[n(0)-n\exp\left\{-(\phi_0-V_1)/\boldsymbol{k}T\right\}\right]\\&=F_2\mu\left[n(0)-n\exp\left\{-(\phi_0+V_2)/\boldsymbol{k}T\right\}\right]\end{aligned}\right\}\qquad\ldots\ldots(23)$$

where the usual approximation

$$\int_{-L}^{0}\exp\left(\frac{U}{\boldsymbol{k}T}\right)dx\simeq\frac{\boldsymbol{k}T}{F_1}\exp\left(\frac{\phi_0-V_1}{\boldsymbol{k}T}\right)$$

has been adopted.

This leads to two relations

$$n(0)=n\exp\left\{-\left(\frac{\phi_0-V_1}{\boldsymbol{k}T}\right)\right\}\left[\frac{F_1-F_2\exp\left(-V/\boldsymbol{k}T\right)}{F_1-F_2}\right]$$

$$I=-n\mu\frac{F_1F_2}{F_1-F_2}\exp\left\{-\left(\frac{\phi_0-V_1}{\boldsymbol{k}T}\right)\right\}\left[1-\exp\left(-\frac{V}{\boldsymbol{k}T}\right)\right]$$

which have been obtained by Taylor, Odell and Fan and are used by them to deduce the zero bias resistance.

Elimination of V_1 gives

$$I=-\mu F_1F_2n(0)\frac{1-\exp\left(-V/\boldsymbol{k}T\right)}{F_1-F_2\exp\left(-V/\boldsymbol{k}T\right)}.\qquad\ldots\ldots(24)$$

The condition for a stationary $n_{\rm d}$ now becomes

$$2n(0)\,\boldsymbol{e}\left(\frac{\boldsymbol{k}T}{2\pi\boldsymbol{m}}\right)^{1/2}c\,(N_{\rm d}-n_{\rm d})=\boldsymbol{e}n_{\rm d}/\tau.$$

Comparing with equation (4*a*),

$$\left.\frac{n_{\mathrm{d}}}{N_{\mathrm{d}}}=\frac{\dfrac{\alpha}{1-\alpha}\cdot\dfrac{n(0)}{n\exp(-\phi_0/\boldsymbol{k}T)}}{1+\dfrac{\alpha}{1-\alpha}\,\dfrac{n(0)}{n\exp(-\phi_0/\boldsymbol{k}T)}}.\right\}\quad\ldots\ldots(25)$$

Hence defining $\sigma_0\equiv n\mu\exp(-\phi_0/\boldsymbol{k}T)$ equations (24) and (25) may be used to eliminate $n(0)$

$$\left.\frac{n_{\mathrm{d}}}{N_{\mathrm{d}}}=\frac{\dfrac{\alpha}{1-\alpha}\,\dfrac{I}{2\sigma_0F_0}}{\dfrac{\alpha}{1-\alpha}\,\dfrac{I}{2\sigma_0F_0}-\dfrac{F_1F_2\{1-\exp(-V/\boldsymbol{k}T)\}}{2F_0\{F_1-F_2\exp(-V/\boldsymbol{k}T)\}}}.\right\}\quad\ldots\ldots(26)$$

Finally substituting into equation (20)

$$\left.(\alpha+\beta)\frac{F_1-F_2}{2F_0}=\frac{j'}{j'-\dfrac{F_2}{2F_0}\,\dfrac{1-\exp(-V/\boldsymbol{k}T)}{1-(F_2/F_1)\{\exp(-V/\boldsymbol{k}T)\}}}+\beta\right\}\quad\ldots\ldots(27)$$

where

$$j'=\frac{I}{2\sigma_0F_0}\,\frac{\alpha}{1-\alpha}.$$

When $V\gg\boldsymbol{k}T$ and further

$$-\frac{F_2}{F_1}\exp\left(-\frac{V}{\boldsymbol{k}T}\right)\ll 1\quad\ldots\ldots(28)$$

even though, for a large V, $-F_2\gg F_1$ then equation (27) reduces to

$$(\alpha+\beta)\frac{F_1-F_2}{2F_0}=\frac{j'}{j'-F_2/2F_0}+\beta.\quad\ldots\ldots(29)$$

Comparing with the corresponding equation (15) for the diode theory ($\tanh(V/2\boldsymbol{k}T)\simeq 1$) it will be seen that they are identical if j is replaced by $-j'2F_0/F_2$. Since F_2 increases with V the diffusion theory gives a less rapid increase of I with respect to V than the diode theory, other things being equal.

It can easily be shown that if a quasi-Fermi level is defined for the $n(0)$ electrons by means of the equation

$$n(0)=n\exp\{-(\phi-V_{\mathrm{s}}')/\boldsymbol{k}T\}\quad\ldots\ldots(30)$$

where V_{s}' is again measured from $F_{\mathrm{s}0}$ (figure 2) then the assumption $V_{\mathrm{s}}=V_{\mathrm{s}}'$ (cf. equation (16)) leads to the condition (25). This implies than the $n(0)$ and n_{d} electrons are in thermal equilibrium with one another. This is to be expected in the present case since the use of the diffusion equation implies that at any point x the $n(x)$ electrons are in a slightly perturbed thermal equilibrium with the lattice.

For the sake of simplicity the theory will henceforth be restricted to equation (29) with the corresponding condition (28).

2.3.1. *Modification for a non-symmetrical contact.*

Proceeding as above and with reference to the model of § 2.2.1, the final result is

$$(\alpha+\beta)\frac{F_{\mathrm{a}}-F_{\mathrm{b}}}{F_{\mathrm{a}0}-F_{\mathrm{b}0}}=\frac{j'}{j'-F_{\mathrm{b}}/(F_{\mathrm{a}0}-F_{\mathrm{b}0})}+\beta$$

where

$$j'=\frac{I}{\sigma_0(F_{\mathrm{a}0}-F_{\mathrm{b}}{}^{0})}\,\frac{\alpha}{1-\alpha}.$$

2.4. *Application of Diode Theory to Mott Barriers*

It will now be assumed that there exists a layer (thickness L) on the surface of the semiconductor which has no donors or acceptors for electrons. Then if the charge density of the current carriers can be neglected there will be a constant field in the charge free layer. If $\phi_0 - V_1, \phi_0 + V_2 \gg \boldsymbol{k}T$ such an approximation will be good since the charge carrier densities fall off approximately as $\exp(-U/\boldsymbol{k}T)$ where U is the increase of potential energy in the barrier. There will however be a small space charge barrier at the end of the impurity-free region so that the linearly varying potential in the Mott barrier joins smoothly with the constant potential inside the semiconductor. We will neglect the height of this small space charge barrier but the validity of this should be checked for any particular case.

The reason for introducing the Mott barriers is that they give the correct capacity–voltage variation for the case of untreated SiC contacts (cf. § 2.6).

Considering first the symmetrical case the fields are given by

$$F_0 = K\phi_0/L. \qquad \ldots\ldots(31)$$

By substitution into equation (15) the result can be written in the reduced form

$$v + v_0 + \nu \log\left[1 - \exp\left(\frac{-2v}{\nu}\right)\right] = \frac{j}{\tanh(v/\nu) + 1} + \nu \log j \qquad \ldots\ldots(32)$$

where $v = V(\alpha+\beta)/2\phi_0$; $v_0 = \alpha + \nu\log\{\alpha/2(1-\alpha)\}$; $\nu = \boldsymbol{k}T(\alpha+\beta)/\phi_0$.

If $v \gg \nu$, i.e. $V \gg 2\boldsymbol{k}T$, then $v + v_0 = j/(1+j) + \nu\log j$.

Now the barrier will break down and the current increase very rapidly when $V_1 \to \phi_0$. The value v_c (or V_c) at which this occurs can be found by substituting $V_1 = \phi_0$ into equations (11) and (32). Hence, since $V_c \gg \boldsymbol{k}T$ and $j \gg 1$, $v_c \simeq 1 + \beta$ or $V_c = 2\phi_0(1+\beta)/(\alpha+\beta)$. In fact the theory is inapplicable when $\phi_0 - V_1 \simeq O(\boldsymbol{k}T)$ for then the assumption of charge free barriers is wrong.

For the case of unsymmetrical barriers the final equation can be written in the form (equation (19))

$$v_a + v_{a0} + \nu\log\left[1 - \exp\left\{\frac{v_a}{\nu}\left(1 + \frac{\phi_a'}{\phi_b'}\right)\right\}\right]$$
$$= j_a\left[j_a + \mathrm{Th}\left(\frac{c_b}{c_a}, \frac{v_a}{\nu}\frac{1+(\phi_a'/\phi_b')}{2}\right)\right]^{-1} + \nu\log j_a \qquad \ldots\ldots(33)$$

$$\left.\begin{aligned} \text{where} \qquad v_a &= \frac{\alpha+\beta}{2\phi_a'}V;\ v_{a0} = \alpha + \nu\log\frac{\alpha}{1-\alpha}\frac{c_a}{c_a+c_b} \\ \nu &= \frac{\alpha+\beta}{2}\boldsymbol{k}T\left(\frac{1}{\phi_a'} + \frac{1}{\phi_b'}\right) \end{aligned}\right\} \qquad \ldots\ldots(34)$$

$2\phi_a' = (\phi_{a0}L_b + \phi_{b0}L_a)/L_a$ and similarly for ϕ_b'.

Again, if $v_a \gg \nu$, i.e. $V \gg 2\boldsymbol{k}T$

$$v_a + v_{a0} = \frac{j_a}{1+j_a} + \nu\log j_a \qquad \ldots\ldots(33a)$$

while the breakdown voltage is $v_{ac} \simeq (1-\alpha) + \frac{1}{2}(\alpha+\beta)\phi_{a0}(1/\phi_a' + 1/\phi_b')$

hence $$V_{ac} \simeq \frac{2[(1-\alpha)\phi_a' + \frac{1}{2}(\alpha+\beta)\phi_{a0}(1+\phi_a'/\phi_b')]}{\alpha+\beta}.$$

The first term is usually much larger than the second since $\alpha \ll \beta < 1$. Hence $V_{ac}/V_{bc} \simeq \phi_a'/\phi_b' \simeq L_b/L_a$.

2.5. *Application of the Diffusion Theory to Schottky Barriers*

Instead of an impurity-free surface region a uniform impurity density N will now be assumed. For a sufficiently high value of ϕ_0 (i.e. $\phi_0 \gg \mathbf{k}T$) this leads to the well-known Schottky exhaustion barrier with a variable thickness. The reason for this choice of barrier is that it gives the correct capacity–voltage relation for the Ge grain boundary barrier examined by Taylor, Odell and Fan. The relation between the field and the barrier height is now given by $F_0=(8\pi\mathbf{e}^2N/K)^{1/2}\phi_0^{1/2}$.

Substituting into equation (29)

$$(\alpha+\beta)\left[\frac{(\phi_0-V_1)^{1/2}+(\phi_0+V_2)^{1/2}}{2\phi_0^{1/2}}\right]=\frac{j'}{j'+(\phi_0+V_2)^{1/2}/2\phi_0^{1/2}}+\beta.$$

This can be further simplified if $V\gg\phi_0$ for $\phi_0+V_2=V+(\phi_0-V_1)\simeq V$. Thus if $v'=V(\alpha+\beta)^2/4\phi_0$ then

$$v'^{1/2}=j'\,[j'+v'^{1/2}/(\alpha+\beta)]^{-1}+\beta.$$

The breakdown voltage is given by

$$v_c\simeq(1+\beta)^2 \quad \text{or} \quad V_c\simeq\frac{4(1+\beta)^2}{(\alpha+\beta)^2}\phi_0.$$

For a non-symmetrical barrier the result is

$$v_a'^{1/2}=j'\,[j'+v_a^{1/2}/(\alpha+\beta)]^{-1}+\beta \qquad \ldots\ldots(35)$$

where

$$v_a'=\frac{(\alpha+\beta)^2N_a}{[(N_a\phi_{a0})^{1/2}+(N_b\phi_{b0})^{1/2}]^2}V. \qquad \ldots\ldots(36)$$

Hence for a given current I the voltages applied across the barrier in the two directions should be in the ratio N_a/N_b. The breakdown voltage is given by

$$v_{ac}'=(1+\beta)^2 \quad \text{or} \quad V_{ac}'=\frac{[(N_a\phi_{a0})^{1/2}+(N_b\phi_{b0})^{1/2}]^2(1+\beta)^2}{(\alpha+\beta)^2N_a}$$

whence $V_{ac}/V_{bc}=N_b/N_a$. (This has been pointed out by Taylor, Odell and Fan who derived the result by simply stating that breakdown in each direction will occur for the same value of F_2.)

2.6. *The Capacity of the Barrier*

It is now supposed that with the barrier under a d.c. bias V as described by figure 2 a small amplitude alternating voltage is superimposed. There are two possible extreme cases depending on the frequency $\omega/2\pi$ of the voltage.

(i) If $\omega\tau\gg1$, the electrons on the surface states will retain the equilibrium corresponding to the d.c. bias. At any instant, let the total applied voltage be $V_t=V+\delta V(t)$. The corresponding changes in F_1 and F_2 will be connected by the requirement of a constant n_d, i.e. $\delta F_1=\delta F_2$.

Now the capacity per unit area of the barrier is given by $C=-\mathbf{e}\,\delta Q/\delta V$ where δQ is the charge transferred from crystal 2 to crystal 1 due to the change δV. But $\delta Q=\delta F_1/4\pi\mathbf{e}=\delta F_2/4\pi\mathbf{e}$ so that

$$\frac{1}{C}=-4\pi\left[\frac{\partial V_1}{\partial F_1}+\frac{\partial V_2}{\partial F_2}\right].$$

In the particular case of Mott barriers this becomes

$$\frac{1}{C}=\frac{4\pi}{K}(L_a+L_b) \qquad \ldots\ldots(37)$$

i.e. independent of V.

For Schottky barriers the result remains the same, but now L_a and L_b are functions of V. In fact

$$\frac{1}{C}=\frac{4\pi}{K}\left(\frac{K}{4\pi \boldsymbol{e}^2}\right)^{1/2}\left[\left(\frac{\phi_0-V_1}{N_a}\right)^{1/2}+\left(\frac{\phi_0-V_2}{N_b}\right)^{1/2}\right]$$

which increases with V, i.e. C decreases as V increases.

(ii) If $\omega\tau \ll 1$, the electrons on the surface states will be in equilibrium with V_t at any instant of time. Thus

$$C=\frac{1}{4\pi}\frac{\partial F_2}{\partial V}=\frac{1}{4\pi}\frac{\partial F_2}{\partial V_2}\left(1-\frac{\partial V_1}{\partial V}\right)$$

where $\partial V_1/\partial V$ is calculated from the d.c. relations derived in previous sections. Thus from equation (11)

$$\frac{\partial V_1}{\partial V}=\boldsymbol{k}T\left[\frac{\partial \log I}{\partial V}-\frac{1/\boldsymbol{k}T}{\exp(V/\boldsymbol{k}T)-1}\right]$$

hence

$$C=\frac{1}{4\pi}\frac{\partial F_2}{\partial V_2}\left[\frac{1}{1-\exp(-V/\boldsymbol{k}T)}-\frac{\partial \log I}{\partial(V/\boldsymbol{k}T)}\right].$$

In the particular case of Mott barriers with $V \gg \boldsymbol{k}T$ using equations (31) and (33*a*).

$$C=\frac{K}{4\pi L_b}\left[1-\frac{\nu}{2[\nu+j_a/(1+j_a)^2]}\right],\quad V\gg \boldsymbol{k}T.$$

Hence C increases with increasing j_a and V but at most by a factor 2.

§ 3. Experimental Results

3.1. *Silicon Carbide Crystal–Crystal Contacts*

The above theory will be applied to some very detailed measurements carried out by Parnell (1953) on contacts between a rounded and a flat face each of similar SiC. The untreated contacts, with which we are here concerned, have a capacity which is constant with respect to a d.c. bias of up to 4 volts in each direction, measured in the range 2–15 kc/s. The constancy of the capacity suggests that the barriers are of the Mott type with $\omega\tau$ very much greater than 1. That $\omega\tau \gg 1$ is supported by crude estimates of τ (see below). For want of more exact information the value of K to be used in equation (37) is taken as 7, which corresponds to the refractive index of the sodium D line in transparent SiC. The area of contact has been calculated by Parnell using Hertz's formulae and a Young's modulus $E \simeq 6\times 10^{12}$ c.g.s. units as a reasonable value, i.e. $A=1{\cdot}21\,\pi[PR/E]^{2/3}$ where P is the load and R the radius of curvature of the rounded point determined microscopically. From these measurements the quantity L_a+L_b may be determined according to equation (37) (cf. table).

Parnell's voltage–current characteristics were taken up to 4 volts for the black and 2 volts for green contacts. The complete (forward and reverse) characteristics for one pair of contacts are shown in figure 3. More detailed characteristics were taken in the range 0·01 to 0·2 volt and found to be almost ohmic. One case (black) is shown in figure 4. By comparing the theoretical curve of $\log j$ plotted against v_a+v_0 (according to equation (33*a*)) with the experimental curve of $\log J$ plotted against V it was immediately obvious that the characteristics such as those in figure 3 could be fitted by a proper choice of parameters. Here $J=Aj$ is the total current. The method of fitting was as follows :

The gradient of the theoretical curve at the point of inflection $j_a = 1$ is $(dv_a/d\log j_a)_1 = \frac{1}{4} + \nu$, whence

$$\left(\frac{dV}{d\log J}\right)_1 = \frac{\phi_a'}{2(\alpha+\beta)} + \left(1 + \frac{\phi_a'}{\phi_b'}\right) \boldsymbol{k}T.$$

This is identified with the linear portions on the experimental ($\log J$, V) curves. From the two simultaneous equations the parameters $\phi_a'/(\alpha+\beta)$, $\phi_b'/(\alpha+\beta)$ and ν

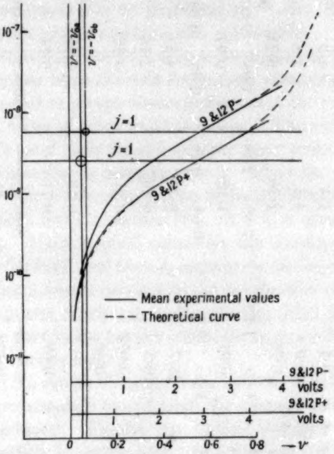


Figure 3. Voltage–current characteristic of a SiC contact.

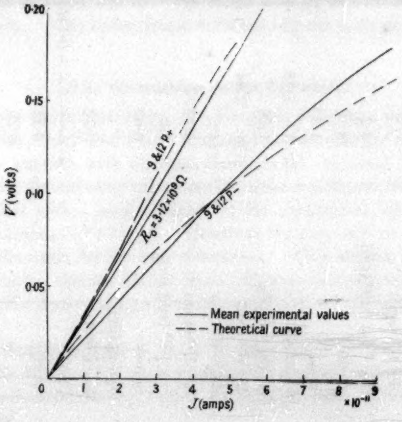


Figure 4. Voltage–current characteristic of a SiC contact for very low voltages.

	← Green					→	← Black					→
	2*b*–5		4–5		3–7		9–12		8–12		9–11	
	$P+$	$P-$	$P+$	$P-$	$P+$	$P-$	$P+$	$P-$	$P+$	$P-$	$P+$	$P-$
$A \times 10^6$ (cm^2)	3·7		1·3		2·5		2·0		2·0		1·8	
$(L_a + L_b) \times 10^6$ (cm)	5·2		2·0		6·1		1·4		1·7		2·0	
$\alpha \times 10^3$	28		16		1·0		4·8		1·5		0·37	
c_a/c_b	1·43		0·51		0·75		2·36		0·25		0·50	
$\phi'/(\alpha+\beta)$	2·25	1·75	1·11	1·20	1·09	1·20	2·59	2·19	1·95	2·02	2·26	2·26
$\phi_a'/\phi_b' = L_b/L_a$	1·28		0·92		0·91		1·18		0·97		1·0	
$L \times 10^6$ (cm)	2·26	2·94	1·05	0·95	3·2	2·9	0·64	0·76	0·86	0·84	1·0	1·0
$AI_0 \times 10^{12}$ (A)	13·8		46·5		25·7		8·20		4·33		0·296	
$I_0 \times 10^6$ (A cm^{-2})	3·72		35·8		10·3		4·10		2·16		0·164	
$c\tau \times 10^{16}$ (cm^2 sec)	6·0		0·36		0·077		0·91		0·56		1·81	
ϵ_d (ev)	0·62		0·55		0·50							
$E_g - \epsilon_a$ (ev)							0·58		0·57		0·59	
$\phi_0 - \zeta$ (ev)	0·72		0·66		0·69		0·71		0·74		0·79	
ζ (ev)	← $-0{\cdot}14(n \simeq 10^{17}$ cm$^{-3})$ →						← $-0{\cdot}20(n \simeq 10^{18}$ cm$^{-3})$ →					
ϕ_0 (ev)	0·58		0·52		0·57		0·51		0·54		0·59	
β	0·29		0·45		0·50		0·21		0·27		0·26	
$\frac{1}{2}N_d \times 10^{-13}$ (cm^{-2})	2·96		0·45		1·45		1·60		1·14		1·10	
$\frac{1}{2}N_a \times 10^{-13}$ (cm^{-2})	3·80		0·65		2·18		1·32		0·90		0·87	
V_c (ev)	5·7	4·7	3·3	3·4	3·3	3·5	6·2	5·4	5·0	5·1	5·7	5·7

could be determined. The experimental curves were then replotted as log J against v_a and v_b respectively. By superimposition of these curves with the theoretical curves using the value of ν thus found (a very good fit being obtained in all cases) the ratios j_a/J, j_b/J and the constants v_{a0}, v_{b0} could be determined. From the former AI_0 and c_b/c_a can be calculated ; using this ratio α can be found from either V_{a0} or v_{b0} (equation (34)). The two values of α so deduced should be the same and their consistency is a check on the theory. Using these parameters the (J, V) characteristic for the voltages which are not much greater than $\boldsymbol{k}T$ was then deduced according to equation (33) by graphical methods. It will be seen from figure 4 that the agreement in this region is also reasonable.

The parameters deduced for the six contacts measured by Parnell are given in the table. The theory has been derived for n-type semiconductors (green) but can be immediately applied to p-type semiconductors if the roles of electrons and donors are replaced by those of holes and acceptors. The subscripts 'a' and 'b' of the theory have been used with the convention that carriers flow from a to b when the pointed SiC member is positive (P+). Thus for green the point is b and for black it is a. The product $c\tau$ can be deduced from equation (4a) and ϵ_d (green) or $E_g - \epsilon_a$ (black) from equation (8). $\phi_0 - \zeta$ is given by equations (4) and (7) (eliminating n). Using typical values for the density of carriers as found by Busch and Labhart (1946) ζ, and then ϕ_0 separately can be estimated. (Silicon carbide probably conducts due to carriers in the conduction or valency bands and also in the excited impurity levels (cf. Baltensberger 1953 and Erginsoy 1952). The latter will not take part in the barrier phenomena since there are supposed to be no impurities in the barrier layer.)

From equation (3a) N_d can be obtained if the value of $K \simeq 7$ is again used.

Since β can be approximately found from the average values of $\phi'/(\alpha+\beta)$ and ϕ_0, N_a can also be deduced. Finally the critical voltages were derived from equation (34) simplified to $V_{ac} \simeq 2\phi_0 + 2\phi_a'/(\alpha+\beta)$ since $\alpha \ll 1$ and the difference between ϕ_{a0} and ϕ_{b0} cannot in any case be found.

The product $c\tau \simeq 10^{-16}\,\text{cm}^2\,\text{sec}$. Now it is unlikely that c will be larger than the cross section of an atom, say $10^{-15}\,\text{cm}^2$, thus leading to a value of τ which is of the order of 10^{-1} sec. This is consistent with the requirement previously adopted that $\omega\tau \gg 1$.

3.2. *Germanium Grain Boundary*

The voltage–current characteristic of a Ge grain boundary barrier is shown in figure 5 after Taylor, Odell and Fan. Capacity measurements indicated the presence of Schottky barriers with unequal densities (cf. equation (37)). Thus the (J, V) characteristic should obey equation (35) since condition (28) is well satisfied. Figure 5 shows that with a proper choice of the parameters agreement between theory and experiment can be found. However the ratio of j' to J and the ratio of v to V were not the same for the two directions. This should not occur for the former pair and implies that the theory omits a further lack of symmetry. We have nevertheless used the theory as a preliminary explanation, taking the average of the ratios j'/J.

It was found that with $\beta \simeq 0$ $V_a/v_a' = 92$ volts and $V_b/v_b' = 170$ volts. Thus from equation (36) $N_a/N_b \simeq 1{\cdot}8$, somewhat higher than the values ($\simeq 1{\cdot}4$) deduced from capacity measurements by Taylor, Odell and Fan. Also if $\phi_{a0} \simeq \phi_{b0} \simeq \phi_0$, then $\phi_0 \simeq 38{\cdot}7\,\alpha^2$ ev.

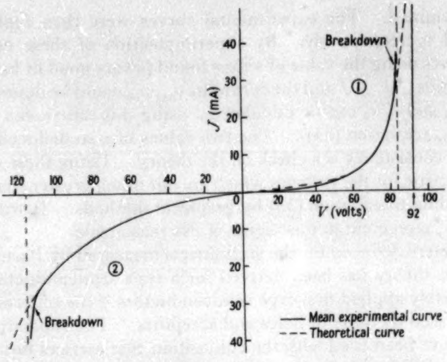


Figure 5. Voltage–current characteristic of a Ge grain boundary barrier.

From measurements on the temperature variation of G_{e0} (the conductivity at zero bias) a value $\phi_0 \simeq 0{\cdot}28$ ev was deduced by Taylor, Odell and Fan. Thus $\alpha \simeq 8{\cdot}5 \times 10^{-2}$.

The ratio $J/\alpha j'$ was found to be 1·5 mA and 8·0 mA for the two directions respectively. Hence taking the average,

$$A(F_{a0} - F_{b0})\sigma_0(1-\alpha)/\alpha^2 \simeq 4{\cdot}25 \times 10^{-3}\ \mathrm{A\,cm^{-2}}.$$

Now $F_{a0} \simeq -F_{b0} \simeq F_0$ and $A = 0{\cdot}019\ \mathrm{cm}^2$ (private communication from Professor Fan) so that $\sigma_0 F_0 \simeq 8{\cdot}1 \times 10^{-4}\ \mathrm{A\,cm^{-2}}$. This product is related to the value G_{e0} of the same temperature by the relation (cf. Taylor, Odell and Fan 1952, eqn (16)) $\sigma_0 F_0 = 2\boldsymbol{k}TG_{e0}/\boldsymbol{e}$ which gives $\sigma_0 F_0 = 7{\cdot}5 \times 10^{-4}\ \mathrm{A\,cm^{-2}}$ if the measured value of G_{e0} is used. The agreement is fortuitously good but supports the validity of the theory.

From equation (3*a*) the density of surface states is $\frac{1}{2}N_d \simeq \frac{1}{2}N_a \simeq 4{\cdot}3 \times 10^{11}\ \mathrm{cm}^{-2}$ where F_0 is calculated from the measured values of L given by Taylor, Odell and Fan. Also from equation (4*a*) $c\tau \simeq 2{\cdot}25 \times 10^{-18}\ \mathrm{cm^2\,sec}$ while $\epsilon_d \simeq 0{\cdot}49$ ev (equation (8)). This value of ϵ_d is too large to satisfy the original requirements, inequality (1). However for the large voltages in question the hole current, which is constant, is in any case negligible as may be seen from equation (25) of Taylor, Odell and Fan (it is less than about 1 mA). Further, the acceptor states will be permanently filled. Thus the theory should apply for the high currents and voltage of figure 5.

From the value of the product $c\tau$, assuming c to be of order $10^{-15}\ \mathrm{cm}^2$, we find τ of order 10^{-3} sec. Since the capacity measurements were carried out in the megacycle range $\omega\tau \gg 1$, justifying the use of equation (37) for the capacity as a function of the applied voltage.

§ 4. DISCUSSION

In figure 6 the voltage–current characteristic of the SiC–SiC contact 9 and 12 has been plotted in a form appropriate to commercial, non-ohmic, bonded silicon carbide grit resistors. The grit size was taken as given by the radius of curvature of the pointed crystal ($\simeq 0{\cdot}0072$ cm, i.e. approximately '100' grit on the standard scale). It will be seen that the region marked as B (barrier region) has the typical

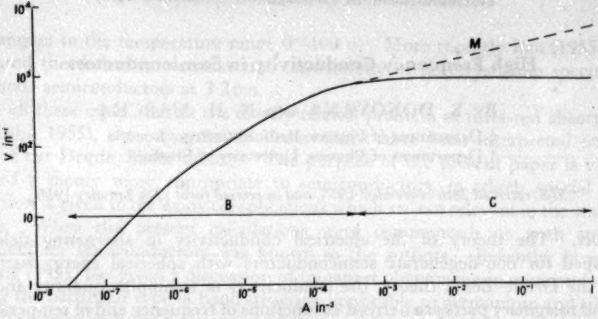


Figure 6. Voltage–current characteristic for an array of SiC point contacts similar to 9 and 12.

shape of the silicon carbide grit resistor characteristic (Mitchell and Sillars 1949) and corresponds to the region of the single contact measurements. However after a voltage gradient of about 700 v in^{-1} has been reached the rapid increase of current (region C) due to the almost complete suppression of the barrier is not observed. Instead the experimental curve follows a line *M* with a simple power law relation. This implies that when the barrier becomes ineffective another mechanism takes over to impede the current flow. Two possibilities present themselves and there may be others. Firstly, the assumption that the potential drop $\phi_0 + V_2$ (cf. figure 2) occurs across the Mott barrier of thickness *L* will break down for high values of *V*. This is because the space charge region at the end of the barrier becomes increasingly important as the field F_2 rises so that only part of the potential drop will occur across the Mott barrier. Secondly, there may be further surface levels above the D states with smaller activation energies. Now the rapid increase of current occurs when the D level is just below the Fermi level of crystal 'a' since it is then almost full. Thus the presence of these higher surface states would give further possibilities for charging the contact surface leading to higher breakdown voltages. No estimates of these effects have so far been made, due to the lack of information about the impurity distribution outside the Mott barrier and the nature of the surface states.

ACKNOWLEDGMENTS

The author wishes to thank Professor H. Fröhlich for the interest he has shown in the work, Dr. R. W. Sillars for suggesting the problem and many stimulating discussions, Mr. R. P. Chasmar for advice on the preparation of the manuscript, Dr. Willis Jackson, Director of Research and Education, and Mr. B. G. Churcher, Manager of the Research Department, Metropolitan-Vickers Electrical Co. Ltd., for permission to publish this paper.

REFERENCES

BALTENSBERGER, W., 1953, *Phil. Mag.*, **44**, 1355.
BUSCH, G., and LABHART, H., 1946, *Helv. Phys. Acta*, **19**, 167.
ERGINSOY, H. C., 1952, *Ph.D. Thesis*, University of London.
MITCHELL, E. W. J., and SILLARS, R. W., 1949, *Proc. Phys. Soc.* B, **62**, 509.
PARNELL, T. M., 1953, *Ph.D. Thesis*, University of London.
TAYLOR, W. E., ODELL, N. H., and FAN, H. Y., 1952, *Phys. Rev.*, **88**, 867.

High Frequency Conductivity in Semiconductors

By B. DONOVAN † and N. H. MARCH ‡

† Department of Physics, Bedford College, London
‡ Department of Physics, University of Sheffield

MS. received 28th November 1955, and in revised form 13th January 1956

Abstract. The theory of the electrical conductivity in alternating fields is developed for non-degenerate semiconductors with spherical energy surfaces. As in the Drude–Zener theory, the conductivity is a complex quantity and the real and imaginary parts are derived as functions of frequency and of temperature.

The cases of lattice scattering and impurity scattering are investigated separately for single-band models and the results are applied to intrinsic semiconductors by considering the two-band model with lattice scattering, and to impurity semiconductors by combining the scattering mechanisms. The loss tangent is also calculated and a comparison is made with the available experimental data.

§ 1. INTRODUCTION

The fundamental ideas in the theory of high frequency conductivity in solids were developed originally by Drude and Lorentz in their treatment of optical dispersion in absorbing media. In the case of metals, the simple free-electron model enables the behaviour in periodic fields to be interpreted in terms of oscillators of zero frequency. As a consequence of a finite damping constant the conductivity (defined as the ratio of current density to electric field strength) is a complex quantity, of which the real and imaginary parts are associated respectively with the measured conductivity and dielectric constant. This model was used by Zener (1933) to explain the transparency of the alkali metals in the ultra-violet.

Following more exact theories of the solid state, quantum mechanical treatments of dispersion in metals were developed by Kronig (1929, 1931), Wilson (1935) and others. For the case of 'quasi-free' electrons, in which the influence of the periodic field of the lattice is small, the high frequency behaviour may be described by expressions identical in form to those of Drude. In the later version, however, the damping constant may be identified with the reciprocal of the relaxation time of the conduction electrons.

These essentially classical expressions, referred to frequently in the literature as the Drude–Zener treatment, have been widely applied to high-frequency phenomena in metals and also, in the absence of any alternative, to the relevant experiments on semiconductors. Thus Benedict and Shockley (1953) measured the dielectric constant of n-type germanium at 1·24 cm over a temperature range 160°–300°K and deduced values for the effective mass and relaxation time of the electrons. Corresponding results for the holes in p-type germanium were subsequently reported by Benedict (1953), and similar measurements were made by Goldey and Brown (1955). Further experiments on germanium at 1·24 cm were carried out by Klinger (1953), who investigated the conductivity and loss-tangent in the temperature range 0°–100°c. More recently Rau (1955) has measured the dielectric constant and conductivity of polycrystalline specimens of various semiconductors at 3·2 cm.

In all these cases, and in the closely related problem of infra-red absorption (cf. Kahn 1955), the experimental observations have been interpreted on the basis of the Drude–Zener theory. The purpose of the present paper is to put forward a theory more appropriate to semiconductors, in which special consideration is given to the many complications which are not relevant in the study of metals. When the present calculations were commenced, no such specific treatment for semiconductors was known to the authors. Recently, however, Herring (1955) has given a discussion of transport phenomena in semiconductors having the particular type of band structure applicable to germanium and silicon. This discussion includes a derivation of the complex conductivity with particular reference to the experiments of Benedict and Shockley.

In the present paper the emphasis is placed on different aspects of the problem. The simplest possible band structure (quasi-free electrons) is adopted and the significance of the various scattering mechanisms is considered.

§ 2. OUTLINE OF METHOD AND RELATION TO DRUDE–ZENER EQUATIONS

The essential problem is the evaluation of the complex conductivity $\sigma_c(\omega)$ which we shall write as

$$\sigma_c(\omega)=\sigma_0(\alpha+i\beta)=\sigma_r+i\sigma_x, \qquad \ldots\ldots(1)$$

where σ_0 is the d.c. conductivity and ω the angular frequency of the applied field.

The physically significant quantities are the measured conductivity σ_r and the dielectric constant $\kappa(\omega)$. If the dielectric constant of the lattice in the absence of free carriers is denoted by κ_0 we then have

$$\kappa_0-\kappa(\omega)=-\frac{4\pi}{\omega}\sigma_x. \qquad \ldots\ldots(2)$$

For the sake of comparison we quote the well-known Drude–Zener expressions for a system of carriers with relaxation time τ and effective mass m. If n is the carrier density, we have†

$$\left.\begin{aligned}\sigma_c(\omega)&=\frac{\sigma_0}{1+i\omega\tau}, &\quad \sigma_0&=\frac{ne^2\tau}{m};\\ \alpha&=\frac{1}{1+\omega^2\tau^2}, &\quad \beta&=-\frac{\omega\tau}{1+\omega^2\tau^2}.\end{aligned}\right\} \qquad \ldots\ldots(3)$$

These expressions apply to a single band model in which the carriers are treated as a degenerate gas; the frequency dependence is obtained by replacing $1/\tau$ by $(1+i\omega\tau)/\tau$ in the usual formulation, as is clear from the Boltzmann equation. In semiconductors we have essentially simply to average (3) over the Maxwellian velocity distribution, taking account of the variation of the relaxation time τ with velocity. Following the usual calculation (cf. Wilson 1953, p. 231) the appropriate equation giving the complex conductivity is

$$\sigma_c(\omega)=\frac{4\pi ne^2}{3kT}\left(\frac{m}{2\pi kT}\right)^{3/2}\int_0^\infty\frac{\tau v^4}{1+i\omega\tau}\exp\left(-\frac{mv^2}{2kT}\right)dv. \qquad \ldots\ldots(4)$$

As a variable analogous to $\omega\tau$ we choose

$$u=\frac{3\pi^{1/2}}{4\rho_0}\left(\frac{\omega m}{ne^2}\right) \qquad \ldots\ldots(5)$$

---

† Gaussian units are used throughout this paper.

impurity scattering is negligible in an intrinsic semiconductor, we can extend the preceding treatment to the two-band model and we distinguish the electrons and holes by suffixes 1 and 2 respectively.

Since the two bands contribute independently to the current, the complex conductivity may be obtained by a straightforward extension of the method in § 3. The general expression is

$$\sigma_c(\omega)=\frac{4e^2}{3(2\pi kT)^{1/2}}\left\{\frac{n_1 l_1}{m_1^{1/2}}(\alpha_1+i\beta_1)+\frac{n_2 l_2}{m_2^{1/2}}(\alpha_2+i\beta_2)\right\} \quad \ldots\ldots(11)$$

where α_1 and α_2 denote expression (8) evaluated with u_1 and u_2 respectively, and similarly for β_1 and β_2. At sufficiently high temperatures the concentrations of electrons and holes may be regarded as equal, and we therefore write $n_1=n_2=n$. If, further, we introduce the ratios

$$l_1/l_2=\lambda, \qquad m_1/m_2=\nu,$$

we have $u_1/u_2=\lambda\nu^{1/2}$ and the ratio of electron mobility μ_1 to hole mobility μ_2 is given by

$$b=\mu_1/\mu_2=l_1 m_2^{1/2}/l_2 m_1^{1/2}=\lambda/\nu^{1/2}. \quad \ldots\ldots(12)$$

With this notation (11) becomes

$$\sigma_c(\omega)=ne\mu_2\{b(\alpha_1+i\beta_1)+(\alpha_2+i\beta_2)\}$$

and since the d.c. conductivity is $\sigma_0=ne\mu_2(b+1)$ the complex conductivity may be expressed as $\sigma_c(\omega)=\sigma_0(\alpha+i\beta)$,

$$\text{where} \quad \alpha=\frac{b\alpha_1+\alpha_2}{b+1} \quad \ldots\ldots(13) \quad \text{and} \quad \beta=\frac{b\beta_1+\beta_2}{b+1}. \quad \ldots\ldots(14)$$

Figure 3 shows α plotted against u_1 for four different combinations of the parameter λ and ν. The values 1 and 10 have been taken for λ and two values of ν have been considered with each value of λ. These values of ν have been chosen to give particular values for the mobility ratio b; for curves *a* and *c*, $b=3$, which is typical of elemental semiconductors, and for curves *b* and *d*, $b=50$, representative of the considerably larger order of magnitude observed in intermetallic compounds.

All these curves are similar in form to figure 1, curve *a*, which, if plotted in figure 3, would lie between curves *b* and *d*. The fact that these three curves are very nearly coincident is to be anticipated, since the high mobility ratio (combined with equal carrier densities) signifies that the current is carried almost entirely by one band. With lower mobility ratios a greater departure from single band behaviour is possible, depending upon the extent to which u_1 differs from u_2. This is illustrated by curve *c*, for which $u_1/u_2 \gg 1$.

Corresponding results for the imaginary part of the complex conductivity are depicted in figure 4, in which β is plotted against u_1.

§ 5. SINGLE BAND MODEL WITH IMPURITY SCATTERING

We now take up the case of an impurity semiconductor, in which conduction is due almost entirely to electrons supplied by impurity donor levels or to holes in the valence band arising from excitation of electrons to impurity acceptor levels. For this type of semiconductor it is satisfactory to consider only a single energy band since intrinsic conductivity may be regarded as negligible. However, an additional feature is encountered here in connection with the scattering of the

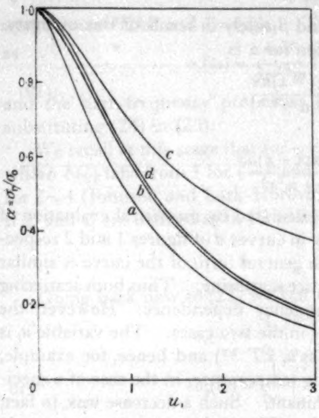


Figure 3. Variation of $\alpha=\sigma_r/\sigma_0$ with parameter u_1 for two-band model: *a*, with $\lambda=1$, $\nu^{1/2}=1/3$; *b*, with $\lambda=1$, $\nu^{1/2}=1/50$; *c*, with $\lambda=10$, $\nu^{1/2}=10/3$; *d*, with $\lambda=10$, $\nu^{1/2}=1/5$. Curve for single band ($\lambda=\nu=1$) lies between *b* and *d*.

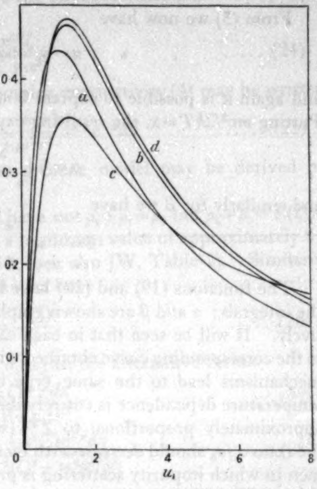


Figure 4. Variation of $\beta=\sigma_x/\sigma_0$ with parameter u_1 for two-band model : *a*, with $\lambda=1$, $\nu^{1/2}=1/3$; *b*, with $\lambda=1$, $\nu^{1/2}=1/50$; *c*, with $\lambda=10$, $\nu^{1/2}=10/3$; *d*, with $\lambda=10$, $\nu^{1/2}=1/5$.

conduction electrons; to account for the observed temperature dependence of conductivity it is necessary to include, in addition to the lattice scattering already discussed, the scattering by the impurity ions.

In this section we consider a model in which the scattering is due to the latter process only and the following calculations are based on the treatment of Conwell and Weisskopf (1950). This involves the essential ideas of Rutherford scattering and leads to a mean free path proportional to the square of the electronic energy. Thus for the relaxation time we write

$$\tau=av^3, \qquad \ldots\ldots(15)$$

with

$$a=\frac{\kappa_0{}^2m^2}{2\pi Ne^4\log\{1+36\kappa_0{}^2d^2k^2T^2/e^4\}},$$

where N is the number of impurity ions per unit volume, $2d$ their average separation and κ_0 the static dielectric constant as before.

The complex conductivity in this case is obtained by substituting (15) in (4), which leads to

$$\sigma_c(\omega)=\frac{\sigma_0}{3}\left(\frac{m}{2kT}\right)^4\int_0^\infty\left\{\frac{v^7}{1+a^2\omega^2v^6}-i\frac{a\omega v^{10}}{1+a^2\omega^2v^6}\right\}\exp\left(-\frac{mv^2}{2kT}\right)dv \quad \ldots\ldots(16)$$

where σ_0 is the d.c. conductivity given by (cf. Wilson 1953, p.8)

$$\sigma_0=\frac{8ane^2}{\pi^{1/2}m}\left(\frac{2kT}{m}\right)^{3/2}. \qquad \ldots\ldots(17)$$

From (5) we now have

$$u = u_i = 6a\omega\left(\frac{2kT}{m}\right)^{3/2} \qquad \ldots\ldots(18)$$

and again it is possible to express both α and β solely in terms of this quantity. Putting $mv^2/2kT = x$, the resulting expression for α is

$$\alpha = 6\int_0^\infty \frac{x^3 \exp(-x)dx}{36 + u_i{}^2x^3}, \qquad \ldots\ldots(19)$$

and similarly for β we have

$$\beta = -u_i\int_0^\infty \frac{x^{9/2}\exp(-x)dx}{36 + u_i{}^2x^3}. \qquad \ldots\ldots(20)$$

The functions (19) and (20) have been calculated by numerical evaluation of the integrals; α and β are shown graphically in curves *d* of figures 1 and 2 respectively. It will be seen that in each case the general form of the curve is similar to the corresponding curve obtained with lattice scattering. Thus both scattering mechanisms lead to the same type of frequency dependence. However, the temperature dependence is entirely different in the two cases. The variable u_i is approximately proportional to $T^{3/2}$ (whereas $u_t \propto T^{-3/2}$) and hence, for example, the ratio σ_r/σ_0 should *decrease* with increasing temperature, in the case of a specimen in which impurity scattering is predominant. Such a decrease was, in fact, observed by Klinger for a germanium crystal but if the specimen was nearly intrinsic, as the measurement of d.c. conductivity would seem to suggest, then an interpretation of the results would not be expected in terms of the model considered in this section. Further discussion of this difficulty is given in § 7.

§ 6. SINGLE BAND MODEL WITH COMBINED SCATTERING MECHANISMS

Having dealt with two models in which the scattering of the electrons is due separately to the lattice vibrations and to impurity ions, we finally consider a more realistic approximation to an impurity semiconductor in which these two scattering mechanisms are combined. This treatment is based on the assumption that the two processes are independent and the effective mean free path l is accordingly given by

$$\frac{1}{l} = \frac{1}{l_t} + \frac{1}{l_i} \qquad \ldots\ldots(21)$$

where l_t and l_i are the mean free paths associated with thermal (lattice) scattering and impurity scattering repectively. Equation (21) has been applied frequently in the analysis of conductivity data and Hall coefficients (see, for example, Johnson and Lark-Horovitz 1950, 1951, Jones 1951).

It is convenient to express the individual mean free paths in terms of the individual d.c. resistivities. Thus if ρ_t is the resistivity when only thermal scattering is operative we have

$$l_t = \frac{3(2\pi mkT)^{1/2}}{4ne^2\rho_t} \qquad \ldots\ldots(22)$$

and if ρ_i is the resistivity when only impurity scattering is operative we have

$$l_i = \frac{\pi^{1/2}m^{5/2}v^4}{(2kT)^{3/2}8ne^2\rho}. \qquad \ldots\ldots(23)$$

Combination of (21), (22) and (23) leads to the effective relaxation time $\tau(=l/v)$ in the form

$$\tau=\frac{\pi^{1/2}mx^{3/2}\zeta}{8ne^2\rho_\mathrm{i}(\zeta+x^2)} \qquad \ldots\ldots(24)$$

where $x=mv^2/2kT$ and $\zeta=6\rho_\mathrm{i}/\rho_\mathrm{t}$. The complex conductivity (4) may be written as

$$\sigma_\mathrm{c}(\omega)=\frac{4ne^2}{3\pi^{1/2}m}\int_0^\infty\frac{\tau(1-i\omega\tau)}{1+\omega^2\tau^2}\,x^{3/2}\,\mathrm{e}^{-x}dx \qquad \ldots\ldots(25)$$

and the high-frequency properties of the present model may be derived by substituting (24) in (25).

We recall at this stage that for $\omega=0$ we have not $\rho_\mathrm{t}+\rho_\mathrm{i}=\rho_0$ but $\rho_\mathrm{t}+\rho_\mathrm{i}=F(\zeta)\rho_0$, where $F(\zeta)$ falls from 1 for $\zeta=0$ and ∞ to a minimum value of approximately 0·7 for $\zeta\sim 4$ (Johnson and Lark-Horovitz 1951; see also JW, Table I). Similarly, if u_t and u_i are defined as in § 3 and § 5, we now have

$$\frac{1}{u_\mathrm{t}}+\frac{1}{u_\mathrm{i}}=\frac{F(\zeta)}{u}. \qquad \ldots\ldots(26)$$

Going back now to (25) we can write $\sigma_\mathrm{r}(\omega)$ in the alternative forms

$$\sigma_\mathrm{r}(\omega)=\frac{1}{\rho_\mathrm{t}}\int_0^\infty\frac{x^3(x^2+\zeta)\mathrm{e}^{-x}dx}{(x^2+\zeta)^2+u_\mathrm{t}{}^2x^3} \qquad \ldots\ldots(27)$$

or

$$\sigma_\mathrm{r}(\omega)=\frac{1}{6\rho_\mathrm{i}}\int_0^\infty\frac{x^3(1+x^2/\zeta)\mathrm{e}^{-x}dx}{(1+x^2/\zeta)^2+u_\mathrm{i}{}^2x^3/36}. \qquad \ldots\ldots(28)$$

It is now easily shown that (27) reduces to the previous expression for $\sigma_\mathrm{r}(\omega)$ for lattice scattering (see § 3) when we put $\zeta=0$, and (28) obviously reduces to (19) when we put $1/\zeta=0$, corresponding to impurity scattering.

The integral in (27) has previously appeared in a study of the magneto-resistance effect in impurity semiconductors by JW. In this paper the integral is denoted by $\mathrm{J}_1(\beta,\gamma)$, where β and γ correspond to our ζ and $u_\mathrm{t}{}^2$ respectively, and is tabulated for six values of β, including zero, over a range of γ from zero to 4·00. Moreover, $\mathrm{J}_1(0,\gamma)$ is identical with the integral $\mathrm{K}_1(\gamma)$ defined by JW (cf. our previous remarks in § 3) who appear to have overlooked this fact. The same remark applies to integrals $\mathrm{K}_2(\gamma)$ and $\mathrm{J}_2(\beta,\gamma)$.

The present investigation has revealed various errors in Tables I and IV of JW, of which the most serious correspond to $\beta=0$. In addition to minor computing errors, it appears that several values of $\mathrm{J}_1(0,\gamma)$ in Table I have been entered against the wrong values of γ. The accompanying table gives the correct values for all arguments listed by JW.

Table of $\mathrm{J}_1(0,\gamma)(\equiv\mathrm{K}_1(\gamma))$ defined by Johnson and Whitesell

γ	$\mathrm{J}_1(0,\gamma)$	γ	$\mathrm{J}_1(0,\gamma)$
0·0025	0·99753	0·35	0·78806
0·01	0·99041	0·50	0·73073
0·04	0·96447	0·70	0·66881
0·05	0·95649	1·00	0·59635
0·10	0·92015	2·00	0·44531
0·15	0·88828	4·00	0·30153
0·20	0·85973	5·00	0·26056

We have evaluated the integral in equation (27) numerically for a number of values of ζ, and the results thus obtained, together with the tables of JW, have enabled us to plot the sequence of curves shown in figure 1 for α against the variable u given by (5) or (26). The curves plotted in this manner are seen to be similar in form but the temperature dependence, which is implicit in u, will be different in each case. In fact it will be seen from (5) that u is proportional to the carrier mobility μ, hence figure 1 exhibits the temperature dependence of σ_r/σ_0 for any sample for which the (μ, T) relation is known. The magnitude and indeed the sign of the temperature variation is thus determined by the proportion of impurity scattering and although ζ is itself a function of temperature the influence of ζ, as evidenced by figure 1, is not sufficiently great to affect this conclusion.

Turning now to the imaginary part of (25), we can likewise express $\sigma_x(\omega)$ in the alternative forms

$$\sigma_x(\omega) = -\frac{u_t}{\rho_t}\int_0^\infty \frac{x^{9/2}e^{-x}dx}{(x^2+\zeta)^2+u_t^2x^3} \qquad \ldots\ldots(29)$$

or

$$\sigma_x(\omega) = -\frac{u_i}{\rho_i}\int_0^\infty \frac{x^{9/2}e^{-x}dx}{36(1+x^2/\zeta)^2+u_i^2x^3}\,. \qquad \ldots\ldots(30)$$

These are again easily shown to reduce in the appropriate limits to the previous results of §§ 3 and 5. The integral in (29) appeared also in JW; it is identical with $J_2(\beta,\gamma)$.

Carrying out various additional numerical integrations we have obtained the results for σ_x/σ_0 as a function of u shown in figure 2. As before, a compact series of curves is obtained, all having the same general form, with the limiting cases represented by the results of §§ 3 and 5.

§ 7. COMPARISON WITH EXPERIMENT

The relevant experimental data consist in general of measurements of conductivity and dielectric constant, in the microwave region, over a range of temperature. In order that the present theory may be applied to such results the d.c. conductivity must be known over the same range of temperature and this information is given only in the case of Klinger's experiments. The conductivity of a near-intrinsic sample of germanium ($\rho_0 = 56$ ohm cm at room temperature) was measured at 1·24 cm and, as noted previously, the variation of σ_r/σ_0 with temperature was contrary to that predicted by the present theory for the case of predominant lattice scattering.

Moreover, this temperature dependence differs widely from any behaviour within the scope of the present calculations. Thus according to Klinger the values of σ_r/σ_0 at 20° and 100°C are 0·87 and 0·43 respectively, whereas even if we use curve *d* of figure 1, which is the most favourable to give a temperature dependence of the type reported by Klinger, we find a much smaller variation with temperature. To obtain a temperature dependence as marked as that shown in Klinger's experiments it would seem essential to assume an entirely different scattering mechanism. Some calculations we have made indicate that it would be necessary for the mobility to vary roughly as T^5 to account for Klinger's results. A strict quantitative comparison for germanium would require the many-valley model discussed by Herring (1955) but even if this modification is included it is found to be insufficient to cause a substantial change in the above argument.

Klinger has also measured the dielectric constant and from it has deduced the loss tangent, which in our notation is given by

$$\tan\delta = \frac{\alpha}{\beta+\omega\kappa_0/4\pi\sigma_0}. \qquad \ldots\ldots(31)$$

It is of interest to consider the variation of $\tan\delta$ with temperature, as given by the present theory, for representative germanium specimens at the same frequency ($2{\cdot}4 \times 10^{10}$ c/s) and over the same temperature range (20°–100°c). We put $\kappa_0 = 16$ and use the mobility data given by Conwell (1952).

We take first the case of a sample with sufficient impurity content for the mobility to be approximately *constant* in the quoted temperature interval (cf. Sample 58 in Conwell's paper). The low resistivity ($\sim 10^{-2}$–10^{-3} ohm cm) means that $\omega\kappa_0/4\pi\sigma_0 \ll \beta$ and hence $\tan\delta$ is virtually independent of temperature. For such a sample we should have $u \sim 0{\cdot}2$ and hence, from figure 1, curve *d*, $\tan\delta \sim 3{\cdot}7$.

Secondly, we consider an intrinsic sample in which the mobility follows the usual $T^{-3/2}$ variation; the resistivity values are taken from Klinger's graph. If the effective mass is known u can be evaluated as a function of temperature from (5). Figure 5 shows the results obtained with $m^* = m$ and with $m^* = m/4$ (cf. Debye and Conwell 1954), together with the experimental curve found by Klinger. It will be seen that in this instance a change in the effective mass causes a relatively slight modification to the curve. The essential point is that in this case the most influential term in (31) is that containing σ_0; according to our calculations α is sensibly constant in this temperature range, whereas according to Klinger's results it decreases markedly as the temperature increases.

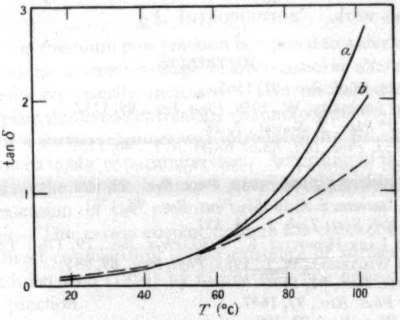


Figure 5. Variation of loss tangent with temperature at 1·24 cm for typical intrinsic specimen of germanium : *a*, with $m^* = m$; *b*, with $m^* = m/4$. Klinger's experimental results are represented by the broken curve.

§ 8. CONCLUSION

In this paper the theory of high-frequency conductivity in semiconductors has been developed on the basis of a model with spherical energy surfaces. The object has been to derive the real and imaginary parts of the complex conductivity in terms of frequency and temperature, and the calculations have been carried through using classical statistics. The treatment is applicable to intrinsic and impurity semiconductors and the results can be used to interpret the observed behaviour of the conductivity and dielectric constant.

The frequency dependence of the real and imaginary parts of $\sigma_0(\omega)$ is found in all cases to be similar to that given by the Drude–Zener theory. The variation with temperature depends on the ratio of impurity scattering to lattice scattering and can be obtained for any specimen if the mobility is given as a function of temperature. The influence of the scattering mechanism is, in fact, the main feature of the present analysis and, although the results for the many-valley model would in general be slightly different, approximate calculations based on Herring's work indicate that our conclusions would not be seriously affected. One further point may be made, arising from figures 1 and 2. The compact series of curves obtained in each case shows that to a fair approximation α and β may be represented in terms of a single parameter, and explains the fact that it has been possible to use the Drude–Zener theory (which contains only one parameter, $\omega\tau$) for semiconductors with any type of scattering.

Finally, the application of the foregoing theory to optical problems should be mentioned. It is possible to derive, for example, the absorption coefficient and hence to investigate the infra-red absorption due to free carriers. Calculations along these lines are now being carried out and the results will be presented in a subsequent paper.

ACKNOWLEDGMENT

We should like to thank a referee, both for drawing our attention to Bronstein's paper and for suggestions which have helped to improve the presentation of this work.

REFERENCES

BENEDICT, T. S., 1953, *Phys. Rev.*, **91,** 1565.
BENEDICT, T. S., and SHOCKLEY, W., 1953, *Phys. Rev.*, **89,** 1152.
BRONSTEIN, M., 1932, *Phys. Z. Sowjet.*, **2,** 28.
CONWELL, E. M., 1952, *Proc. Inst. Radio Engrs, N.Y.*, **40,** 1327.
CONWELL, E. M., and WEISSKOPF, V., 1950, *Phys. Rev.*, **77,** 388.
DEBYE, P. P., and CONWELL, E. M., 1954, *Phys. Rev.*, **93,** 693.
GOLDEY, J. M., and BROWN, S. C., 1955, *Phys. Rev.*, **98,** 1761.
HERRING, C., 1955, *Bell Syst. Tech. J.*, **34,** 237.
JOHNSON, V. A., and LARK-HOROVITZ, K., 1950, *Phys. Rev.*, **79,** 176 ; 1951, *Ibid.*, **82,** 977.
JOHNSON, V. A., and WHITESELL, W. J., 1953, *Phys. Rev.*, **89,** 941.
JONES, H., 1951, *Phys. Rev.*, **81,** 149.
KAHN, A. H., 1955, *Phys. Rev.*, **97,** 1647.
KLINGER, Y., 1953, *Phys. Rev.*, **92,** 509.
KRONIG, R. DE L., 1929, *Proc. Roy. Soc.* A, **124,** 409 ; 1931, *Ibid.*, **133,** 245.
RAU, R. R., 1955, *University of Pennsylvania Technical Report.*
SHIBUYA, M., 1954, *Physica*, **20,** 971.
WILSON, A. H., 1935, *Proc. Roy. Soc.* A, **151,** 274 ; 1953, *The Theory of Metals* (Cambridge : University Press).
ZENER, C., 1933, *Nature, Lond.*, **132,** 968.

Observations on the Growth of Excess Current in Germanium p–n Junctions

By A. R. F. PLUMMER

Research Laboratories of The General Electric Company, Ltd., Wembley, Middx.

MS. received 22nd September 1955 *and in revised form* 12*th January* 1956

Abstract. The excess current induced in a germanium p–n junction by water vapour was found to be time dependent, and three separate relaxation times were observed. By taking the time of growth into account the excess current was resolved experimentally into two components having linear and saturation relationships with the applied voltage. The surface structure assumed to be responsible for the saturation component does not develop until after a reverse voltage has been applied to the junction, and when the applied voltage is removed this surface structure disappears rapidly compared with the time of growth.

The possibility that these components correspond to channel and surface leakage currents is discussed.

---

§ 1. Introduction

When a germanium p–n junction is exposed to water vapour, the shape of the reverse current–voltage characteristic is altered and the reverse currents are usually increased. Two mechanisms which have been proposed to explain this excess current are channel conduction and surface leakage. The concept of a surface channel was introduced by Brown (1953) to explain surface phenomena in the n–p–n transistor. According to this view the presence of free charges on the surface induces a space charge within the semiconductor, creating an extension of the junction and thus causing an increase in the reverse current. The excess current flows in the channel or space charge layer (figure 1). Direct confirmation of the existence of surface channels has been obtained by Christensen (1954) by means of a fine light-spot moved over the surface of the junction.

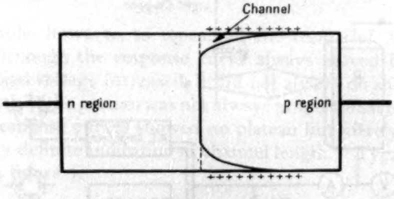


Figure 1.

McWhorter and Kingston (1954) have shown that water vapour can induce a channel on the p-region of a junction and that the excess current is proportional to channel length. Law (1954) has suggested the alternative hypothesis that the excess current is carried by the motion of ions over the surface. In the present work it was found that the excess current could be resolved experimentally into two components which may correspond to channel and surface leakage contributions, and the excess current showed time dependence of a type not previously reported.

§ 2. Experimental Methods

The junction to be studied was placed in a bell-jar through which wet or dry gases could be passed, and the effect of water vapour was studied by observation of the reverse current and by Christensen's light-spot method. The gases used were oxygen and nitrogen (O_2 free) supplied by the British Oxygen Company. The humidity of the gas was controlled by bubbling through water kept at a given temperature, and dry gas was obtained by the use of cold traps. All the experiments were made with the junction at room temperature.

Before each experiment, the junction was given a clean-up etch in CP4 (Haynes and Shockley 1951) followed by thorough rinsing in running water, twice distilled. After removing surplus water with filter paper the junction was placed in the bell-jar and allowed to dry in a current of nitrogen. Both indium alloy junctions (1Ω cm n-type Ge) and grown junctions (5Ω cm n-type, 1Ω cm p-type) were studied.

Reverse currents were measured at constant voltage, an oscilloscope being used to observe rapid changes (figure 2). For the measurement of decay times, as described in § 5, the current was interrupted by means of a rotating commutator having an adjustable gap. In the light-spot technique, a light-spot 1/1000 in. in diameter was moved over the germanium surface and the photovoltage appearing across the junction was plotted against distance moved by the light. This response curve normally shows a sharp peak which marks the position of the junction, but if a channel is present the curve should show a plateau near the junction (Christensen 1954). To permit observation of rapid changes in the response curves, the light spot could also be traversed at 5–10 cycles per second by means of a scanning disc, enabling the curves to be displayed on an oscilloscope.

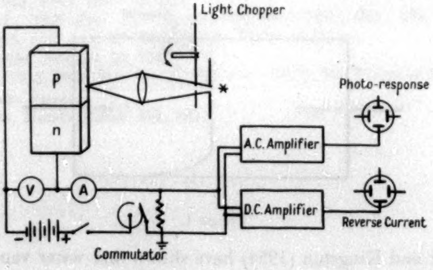


Figure 2. Diagram of apparatus.

§ 3. The Growth of the Excess Current

If a reverse voltage is applied to a p–n junction surrounded by dry gas, the current rises very rapidly to a constant value I_0. In wet gas, however, it was found on applying a voltage that after this initial rise, the current slowly increased to an equilibrium value (figure 3). An increase in applied voltage was always followed by such slow drift. The time of growth depended on the condition of the surface and the humidity, and was largely independent of applied voltage. For the relative humidity used (23%) the time of growth was usually of the order of several minutes. The equilibrium current I_f depended on the condition of the surface, humidity and applied voltage, and often exceeded I_0 by several orders of magnitude.

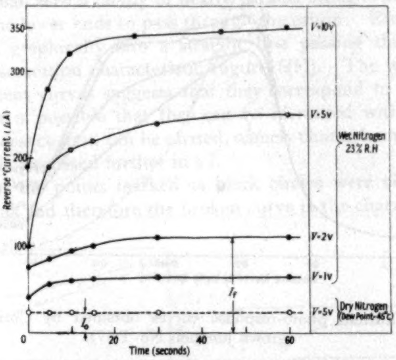


Figure 3. The growth of the excess current (grown junction No. P 17).

3.1. *Light-Spot Observations on Grown Junctions*

The junctions were examined by the light-spot method and it was found that the shape of the photo-response curves gradually changed during the growth of the excess current. As the current increased towards the equilibrium value, the photo-response curve was observed to extend further and further into the p-region.

It was not possible, however, to repeat all the results of McWhorter and Kingston (1954); although the response curve always moved further into the p-region as the applied voltage increased, it did not always do so in the expected manner, and the length of the plateau was not always proportional to excess current.

Very often the response curves showed no plateau but merely became asymmetrical, without any definite indication of channel length. Typical equilibrium results are shown in figure 4.

§ 4. Origin of the Excess Current

If we assume that the excess current flows in some kind of surface structure, then the voltage dependance of the excess current should give information about this surface structure.

It has been shown that the excess current is a function of voltage and time (figure 3), and it will be assumed that it is a function of the state S of the surface. The slow growth of the excess current may then be explained by supposing that the state of the surface changes with time after applying a voltage. Furthermore, the equilibrium state attained depends on the voltage which has been applied (figure 4). Thus S is a function of both voltage and time and may be written $S(V,t)$. Hence $I(V,t) = I_0(V) + I_e[V, S(V,t)]$ where $I(V,t)$ = current measured in the external circuit, $I_0(V)$ = current carried by the junction in dry gas, $I_e(V,S)$ = current flowing in the surface structure.

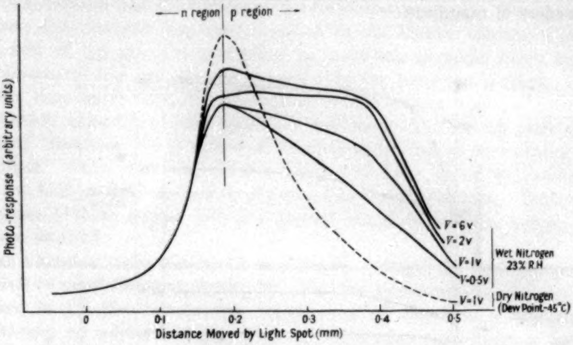


Figure 4. Equilibrium photo-response curves obtained by Christensen's method (grown junction No. P 17).

To study the surface structure we observe $I_e(V,S)$, the current–voltage characteristic of the surface structure, in a given state S. Since $I(V,t)$ and $I_0(V)$ can be measured separately in wet gas and dry gas respectively, this could be done by simply plotting $I(V,t) - I_0(V)$ against V, if it were not for the fact that S is itself a function of V. The surface structure therefore will not remain constant but will be changed by the act of measurement. This difficulty may be avoided if we restrict the measurements to initial values, bearing in mind that S changes relatively slowly with time.

The experimental procedure was as follows: A grown junction was placed in dry gas and the $I_0(V)$ characteristic was measured. The atmosphere was then changed to one of constant relative humidity and a standing voltage V_{st} applied. When the current reached equilibrium, the surface structure attained a certain steady state $S(V_{st})$ and the current–voltage characteristic was now measured in this steady state. At time $t = 0$ a voltage increment ΔV was applied, and the total voltage $V = V_{st} + \Delta V$ and initial current $I(V)_{t=0}$ were measured (figure 5). The voltage increment was then removed, equilibrium restored and the procedure successively repeated for a range of voltage increments. Under these conditions $I(V)_{t=0} = I_0(V) + I_e[V, S(V_{st})]$. Hence by plotting $I(V)_{t=0} - I_0(V)$ against V we obtain the current–voltage characteristic of the surface structure in the state $S(V_{st})$. The experiment was carried out for standing voltages of 0, 1, 2, 5 and 10 volts, with the results shown in figure 6(*a*).

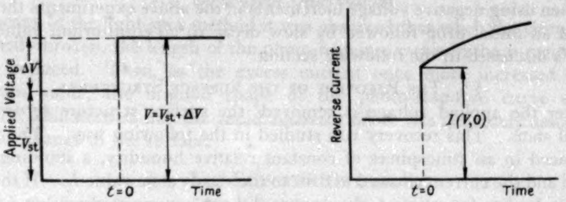
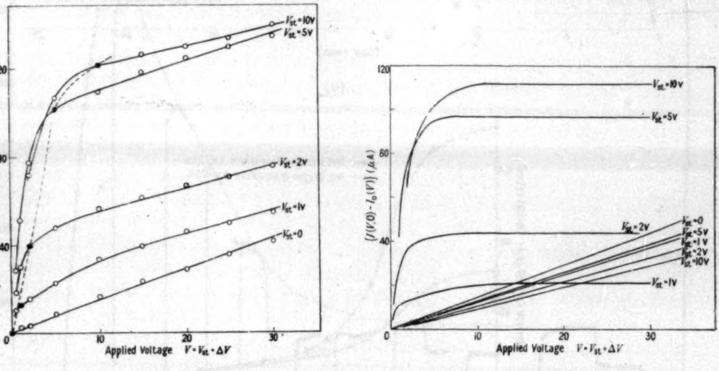


Figure 5.

For V_{st} equal to zero a straight line passing through the origin was obtained. For V_{st} greater than zero a family of nearly parallel straight lines was obtained, which curve at the lower ends to pass through the origin. Each of these curves can be resolved graphically into a straight line passing through the origin, together with a saturation characteristic (figure 6(*b*)). The well-defined shape of these component curves suggests that they correspond to separate physical processes, and it is possible that they can be identified with the two known ways in which excess current can be carried, namely channels and surface leakage. This possibility is discussed further in § 7.

In figure 6(*a*) the points marked as black circles were obtained with zero voltage increments and therefore the broken curve is the characteristic obtained



(*a*) (*b*)

Figure 6 (*a*). Current–voltage Characteristics of the surface structure in different states S (V_{st}) (grown junction No. P 17). Wet nitrogen atmosphere, R.H. = 23%. (*b*) Graphical analysis of the curves of figure 6 (*a*) into linear and saturation components.

by steady state measurements. Several authors (Cutler and Bath 1954, McWhorter and Kingston 1954) have tried to fit a theoretical equation to a characteristic of this kind. The above experiment shows that the state of the surface structure is different for each point on such a characteristic, and the resolution of the excess current at any point into two distinct components may make it easier to carry out a theoretical analysis.

and this caused the excess current to diminish rapidly. Subsequent cycles of wet oxygen and wet nitrogen showed that this change was to some extent reversible, although the excess current never returned to the high value reached just after etching. From these results we conclude that the gradual reduction of the excess current which occurs after etching might possibly be explained by the presence of traces of oxygen in the nitrogen supply, but further work will be necessary to establish this with certainty.

The times of growth and recovery were found to be increased by the presence of oxygen, but the effect was small. For example, figure 7(*b*) shows that the time of recovery for an alloy junction was substantially unaltered by changing the atmosphere from wet nitrogen to wet oxygen, although the excess current was halved.

§ 7. DISCUSSION

The step voltage measurements described in § 4 are essentially pulse measurements, and should give information about the surface structure whilst it is in a certain steady state. We shall assume that during the growth or decay of the excess current, the surface structure is continuously changing from one such state to another. From figure 6 we see that with no standing voltage applied to the junction, the current–voltage characteristic of the surface structure is a straight line passing through the origin, and only after a standing voltage has been applied does a saturation component appear. As the standing voltage is increased, the magnitude of the saturation component increases and the slope of the linear component tends to decrease. Therefore, when a fixed voltage is applied, the initial flow of excess current corresponds to the linear component, and the subsequent growth of the current may be attributed to the growth of the saturation component. We may thus suppose that the rapid decay of the excess current following a momentary interruption of the circuit corresponds to a sudden reduction in the magnitude of the saturation component.

Assuming that the linear and saturation components have physical meaning, we now enquire whether these components can be identified with channel or surface leakage contributions to the excess current. There is nothing in the results to connect the linear component with channel current. On the other hand, we have seen that, as the applied voltage is increased, the length of the photoresponse curve in the p-region and the magnitude of the saturation component both increase. If these changes in the response curve are assumed to indicate the development of a channel on the p-region, then this suggests that the saturation component corresponds to channel current. If this interpretation is correct, it follows that a water-induced channel forms only after a voltage has been applied. Neither a linear nor a saturation relation between channel current and voltage agrees with the theoretical formula given by McWhorter and Kingston (1954), therefore further evidence will be needed before these suggestions can be verified, and the possibility of some other conduction process is not excluded. Clarke (1955) has also obtained results which are not in agreement with McWhorter and Kingston's theory. A leakage current flowing in a surface film might *a priori* be carried either by ions, as suggested by Law (1954), or by electrons. Either of these processes could correspond with the observed linear component.

The rapid recovery effect has not yet been satisfactorily explained.

§ 8. CONCLUSIONS

(i) The excess current induced in a germanium p–n junction by water vapour can be resolved experimentally into two components. They may be related to channel and surface leakage currents, but the possibility of some other conduction process is not excluded.

(ii) The surface structure responsible for one component of the excess current does not develop until after a voltage has been applied to the junction. When the voltage is removed, this surface structure diminishes rapidly compared with the rate of growth.

(iii) The equilibrium value of the excess current slowly diminishes with time. This change is accelerated by exposing the junction to wet oxygen.

ACKNOWLEDGMENT

The author is indebted to the Admiralty for permission to publish this paper.

REFERENCES

BROWN, W. L., 1953, *Phys. Rev.*, **91**, 518.
CHRISTENSEN, H., 1954, *Proc. Inst. Radio Engrs, N.Y.*, **42**, 1371.
CLARKE, E. N., 1955, *Phys. Rev.*, **99**, 1899.
CUTLER, M., and BATH, H., 1954, *J. Appl. Phys.*, **25**, 1440.
HAYNES, J. R., and SHOCKLEY, W., 1951, *Phys. Rev.*, **81**, 838.
LAW, J. T., 1954, *Proc. Inst. Radio Engrs, N.Y.*, **42**, 1367.
McWHORTER, A. L., and KINGSTON, R. H., 1954, *Proc. Inst. Radio Engrs, N.Y.*, **42**, 1376.

On Powder Striations produced by Electric Sparks

By MARY D. WALLER
Formerly at Physics Department, Royal Free Hospital School of Medicine, London

MS. received 24th October 1955, and in final form 16th January 1956

Abstract. The powder striations which may be produced, on either flat surfaces or in tubes, by means of electric sparks are shown to be similar to those obtained in a sound field due to a mechanical vibrator; that is to say they depend upon the presence of aerial vortices. Comparisons of published records show that the frequencies of vibration involved lie in the lower ultrasonic range, and must derive from the shock wave which is associated with the spark. There is no direct relation between the inter-striation distances and the frequencies of the electric circuit.

---

§ 1. HISTORICAL INTRODUCTION

THE first object of the experiments here described was to identify the dust striations which have been formed by means of electric sparks with the dust or powder ridges obtained by the author (Waller 1954, 1955) using a vibrating plate. These in their turn resemble Kundt powder ridges which have been shown by Andrade (1931, 1932) to be due to a hydrodynamic phenomenon in which the sizes of the vortices initiated by the solid particles in a sound field determine the separations of the ridges. The present author's observations with graded powders show how rapidly these inter-ridge distances increase with the diameters of the grains, cf. Bagnold (1941, 1946) where the fundamental part played by the grain diameters is also stressed.

The history of the subject is a long one and begins with Quet and Seguin (1859) who found that the fine particles of carbon powder from lamp black, strewn on a 15 cm × 2 cm glass plate which was situated between the terminals of an electric spark generator, aligned themselves in transverse ridges 2 or 3 mm apart (see also Antolik 1874, 1875). These authors, incidentally, were concerned with trying to obtain effects analogous to the luminous striations in an electric discharge tube.

Mach and his co-workers (1876, 1879) concluded, from experiments with simultaneous sparks, that the carbon dust effects which they obtained on flat surfaces were to be expected from the laws proper to acoustic waves; Rosicky (1876), using elliptical surfaces, was also of this opinion. Gill (1898) believed that both the striated electric discharge and the dust striations in air represented the nodes of standing waves. It is however clear from Cook's (1888) earlier paper that this view is untenable, as the separations between Cook's concentric (lycopodium spore) circular ridges become less as the distance away from the central spark increases. Moreover there is not in his experiments any possibility of the production of stationary waves.

From the papers published in close succession by Richmond (1909), Marsh and Nottage (1911), Robinson (1911) and Barton and Kilby (1912) we learn that lycopodium (the only *uniform* powder employed) produces better results than other powders; that the separation of the ridges increases with the coarseness of the powder (which again precludes a standing wave explanation), and that their distances apart (which vary from 0·3 mm to 1·2 mm in the various records) are unrelated to the high frequencies of the electrical circuits.

The fact that the inter-ridge distances (like those produced by mechanical vibrators) increase with the intensity has also been recorded. Marsh and Nottage, with Richmond's (1909) tube records before them, thought it likely that the striae were produced by a similar mechanism to those in Kundt's tube. But their proof (of what appears now to be a correct surmise) is not convincing, since it depends upon an apparent and partial verification of W. Koenig's (1891) *frictionless* fluid theory of 'attractions' and 'repulsions' between small spheres.

Although Cook thought that the 'sound' of the spark could not account for the phenomenon, Richmond and others have expressed a contrary view. The opinion concerning striations, as Barton and Kilby left it, may be summarized in their words: "that dust figures may be obtained corresponding in their chief features to the ordinary optical phenomena of reflection, refraction, interference and diffraction; but that in addition air currents sometimes acquire circulatory and vortex motion".

Some of the earlier papers mentioned above allude to the striated discharge in gases. This is not our present concern. At the same time it is not without interest to find mention of 'vortices', 'plasticity', 'viscosity', and even 'surface tension', in such papers as those of De la Rue and Mueller (1878, 1883, 1884), Wood (1896), Gassiot (1869). Nor are hydrodynamic analogies (or alternatively mention of nodes and antinodes) entirely lacking from accounts of recent conferences on the electric discharge in gases, see Llewellyn Jones (1952, 1953), von Engel and Francis (1953).

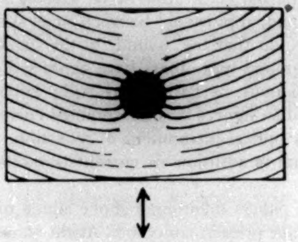
§ 2. FORMATIVE MECHANISM OF ELECTRIC SPARK POWDER STRIATIONS

2.1. *Experiments with Fixed Objects*

Waller (1954, 1955) shows how the presence of various fixed objects affects the alignments of otherwise parallel powder ridges. The results show such a strong resemblance to some of Marsh and Nottage's (1911) figures that we can be left in no doubt that the mechanism when the excitant is a vibrating plate, must be the same as when it is an electric spark. Thus, see the idealized sketch now given, a fixed cylinder modifies the powder pattern in a way which incidentally suggests analogy between hydrodynamic, and magnetic or electrostatic forces. The figure cannot be explained by acoustical laws, but it becomes intelligible in terms of Andrade's (1931, 1932) vortex theory together with the experiments with different sized grains described by the present writer. The small grains of the powder form vortices appropriate to their size. The larger cylinder initiates correspondingly larger vortices which have the effect of removing (or 'repulsing') powder in the direction of the vibration vector and of causing convergence (or 'attraction') of the ridges at right angles to the field.

2.2. *Granular Ridges*

Marsh and Nottage also show a figure of sand grain alignments which are only one particle wide and which may be compared with the present author's so-called *granular* ridges. These are separated by smaller distances than the usual *powder* ridges, the thicknesses and separations of which increase with the amount of powder present. For accurate measurements and for comparison with other 'circulation' observations (e.g. with cylinders or spheres) granular ridges are of special interest. The fact that they have been obtained with the electric spark shows that this method of forming striations is very effective, and strengthens the conclusion that the striae are of the same nature as those obtained in the acoustical experiments mentioned above.



To illustrate how the presence of a fixed cylinder modifies the otherwise parallel powder striations.

§ 3. THE VIBRATION FREQUENCIES

Such evidence as is so far available concerning the frequencies of the periodic disturbance which is responsible for the formation of powder striations by electric sparks is given below.

3.1. *Measurements of Marsh and Nottage's Granular Ridges*

It may be noted that in Marsh and Nottage's (1911) work, the spark from an induction coil was made to pass vertically through a central hole of a powdered glass plate. The authors found that a spark powerful enough to produce an intense condensational pulse was needed to form the ridges. In the discussion afterwards, Dr. Russell stated that "the formation of the striae is one of the best methods of telling whether a spark is oscillatory or not".

Waller has explained that a knowledge of the distance between two ridges, of the diameter of the grains, and of the *ratio* between these two quantities, provides qualitative information concerning the operative sound field vibration frequencies. Taking the mean diameter between the ridges shown in Marsh and Nottage's Fig. 1 as of the order 0·6 mm and the mean 'diameters' of the sand grains as not far removed from 0·3 mm, gives the ratio as about 2. In Waller's records the ratios are greater, they vary from 3 to 8, so that, without suggesting actual frequencies, it can at least be said that the sound frequencies in the electrical experiments must have been much above say 1000 c/s.

3.2. *Richmond's Tube Experiments*

In Richmond's (1909) experiments the spark functioned at one end of a Kundt's tube and the frequencies of the electrical oscillations of the induction coil circuit ranged from about 0·6 to 2 Mc/s. His photographic records show lycopodium striae which, from their close spacing, 0·3 to 0·66 mm, suggest that the frequencies of the sound field are higher than in most Kundt's tube or vibrating plate experiments, i.e. above 3 kc/s. Secondly, there is also a regular variation in the spacing every 3 or 4 mm along the tube suggesting the presence of stationary sound waves of frequencies between about 55 and 44 kc/s.

3.3. *Some Remarks on Experiments at Ultrasonic Frequencies*

3.3.1. *R. Koenig's Kundt's tube records.*

The frequencies mentioned in § 3.1 are ultrasonic. This adds to the interest of re-examining R. Koenig's (1899) photographic records of lycopodium ridges which were obtained in a Kundt's tube, by means of tuning forks, at frequencies ranging from about 4 kc/s up to 90 kc/s. Here, although too much powder is employed for it to be profitable to measure up the inter-ridge distances accurately, we may study how the number of ridges between two nodes gradually diminishes as the frequency increases. At the highest frequency some half dozen ridges can still be recognized in one half wavelength.

3.3.2. *Distinction between the two Kundt repetitive powder phenomena.*

Care is needed when working at high ultrasonic frequencies, not to confuse the two repetitive powder phenomena which may occur in a sounding tube. Campbell, for instance, in his contribution to the discussion on Marsh and Nottage's paper, is not referring to the vortex striae but to the (lycopodium) *nodal streaks* which he and Dye (1911) recorded in a Kundt's tube up to frequencies approaching 1 Mc/s. The sound field was produced by an electrically maintained quartz oscillator. At 900 kc/s the separations are only 0·2 mm which is *less* than separations of any vortex striae so far recorded. Their concluding remark appears significant: "we have in some cases found troublesome effects which require further elucidation, e.g. at frequencies approaching one million only half the number of proper streaks show themselves". This double spacing of 0·4 mm is comparable with the 0·3 to 1·3 mm spacing of the various ridge records mentioned above (§ 1), which suggests that the 'troublesome effects' may possibly be occasioned by interactions between those forces which produce the standing wave, and those which produce the vortex striations.

It is also a matter for some surprise that in books on ultrasonics, e.g. Bergmann (1954), the term 'Kundt dust figure' refers only to the standing wave phenomenon.

§ 4. CONCLUDING REMARKS

The borderland subject dealt with in this paper once more emphasizes the important part which vortex motion often plays in connection with striation phenomena.

At the present day there are many facilities for continuing the research in which, as has already been emphasized, particles of uniform size should be employed. Some of the more recent techniques for obtaining known sonic and ultrasonic frequencies might well be used before proceeding to study the action of the spark itself. This would be produced under standard conditions (see, for instance, Llewellyn Jones 1945).

The spark and the associated shock wave, which is undoubtedly responsible for the production of the striae, is treated by Meek and Craggs (1953), in whose book may be seen a record, taken by G. K. Adams, of sound waves leaving a spark discharge.

REFERENCES

ANDRADE, E. N. DA C., 1931, *Proc. Roy. Soc.* A, **134,** 445; 1932, *Phil. Trans. Roy. Soc.* A, **230,** 417.

ANTOLIK, K., 1874, *Ann. Phys., Lpz.*, **151,** 127; 1875, *Ibid.*, **154,** 14.

BAGNOLD, R. A., 1941, *The Physics of Blown Sand* (London : Methuen); 1946, *Proc. Roy. Soc.* A, **187,** 1.

BARTON, E. H., and KILBY, W. B., 1912, *Phil. Mag.*, **24,** 728.

BERGMANN, L., 1954, *Der Ultraschall*, 6. Auflage (Stuttgart : Hirzel).

CAMPBELL, A., and DYE, D. W., 1911, *Electrician*, **66,** 8620.

COOK, E. H., 1888, *Proc. Phys. Soc.*, **9,** 379.

DE LA RUE, W., and MUELLER, H. H., 1878, *Phil. Trans. Roy. Soc.*, **169,** 155; 1883, *Nature, Lond.*, **28,** 381; 1884, *Ibid.*, **29,** 349.

VON ENGEL, A., and FRANCIS, G., 1953 (Report of International Conference on Ionization Phenomena in Gases, Oxford), *Nature, Lond.*, **172,** 798.

GASSIOT, J. P., 1869, *Rep. Brit. Ass.*, 46.

GILL, A. H. V., 1898, *Amer. J. Sci.*, **55,** 399.

KOENIG, R., 1899, *Ann. Phys., Lpz.*, **69,** 626.

KOENIG, W., 1891, *Ann. Phys., Lpz.*, **42,** 353, 549.

LLEWELLYN JONES, F., 1945, *J. Soc. Chem. Ind., Lond.*, **64,** 317; 1952 (Report of British Association), *Nature, Lond.*, **170,** 601; 1953, *Rep. Progr. Phys.*, **16,** 216, §5, 522 (London : Physical Society).

MACH, E., and GRUSS, G., 1879, *Akad. Wiss. Wien*, **72,** 44.

MACH, E., and WOSKA, J., *Akad. Wiss. Wien*, **78,** 467.

MARSH, S., and NOTTAGE, W. H., 1911, *Proc. Phys. Soc.*, **23,** 264.

MEEK, J. M., and CRAGGS, J. D., 1953, *Electrical Breakdown in Gases* (Oxford : Clarendon Press).

QUET, M., and SEGUIN, M., 1859, *C.R. Acad. Sci., Paris*, **48,** 338.

RICHMOND, T. J., 1909, *Phil. Mag.*, **78,** 771.

ROBINSON, J., 1911, *Phil. Mag.*, **21,** 268.

ROSICKY, W., 1876, *Akad. Wiss. Wien*, **73,** 629.

WALLER, M. D., 1954, *J. Sci. Instrum.*, **31,** 410; 1955, *Proc. Phys. Soc.* B, **68,** 462.

WOOD, R. W., 1896, *Phys. Rev.*, **4,** 191.

Electronic Conduction in Solids with Spherically Symmetric Band Structure

By R. BARRIE

Services Electronics Research Laboratory, Baldock, Herts.

MS. received 15*th December* 1955, *and in final form* 23*rd January* 1956

Abstract. The scattering of electrons whose energy is a function only of the magnitude of their wave number vector is considered for an elastic and an inelastic process. The elastic process is scattering by ionized impurities, the inelastic is scattering by the non-acoustical modes of the lattice vibrations in a homopolar crystal. The times of relaxation for these two processes are calculated.

---

§ 1. INTRODUCTION

IN the treatment of transport phenomena in solids, it is usually assumed that the dependence of the energy of the electron on its wave number vector $\mathbf{k}$ is given by $\epsilon = \hbar^2k^2/2m^*$ where ϵ is the energy of the electron measured from the bottom of the band considered and m^* is its effective mass. It has been found that this standard band gives an insufficient description of the electrons in a large class of solids, and various non-standard bands have been discussed. In this paper, we shall consider bands in which the energy depends only on $|\mathbf{k}|$, but not quadratically. Such a band has been studied by Radcliffe (1955) to explain the properties of pure p-type germanium where the holes are scattered by the acoustical modes of the lattice vibrations; Parmenter (1955) has found a band of this form for electrons in a disordered alloy; Barrie and Edmond (1955) have used such a band to try to correlate the electrical and optical properties of highly degenerate n-type samples of indium antimonide. We shall assume that the band has a minimum at $\mathbf{k}=0$ and that ϵ is monotonically increasing with $|\mathbf{k}|$. If ϵ were not monotonically increasing, the band would be of the many-valley type and one would have to consider inter-valley scattering in the transport phenomena.

§ 2. GENERAL CONSIDERATIONS

The theoretical study of transport phenomena consists of two parts; first, the solution of the Boltzmann transport equation assuming that a time of relaxation τ for the scattering processes exists, with the subsequent derivation of formulae for the various quantities such as conductivity, Hall constant etc.; secondly, the finding of τ as a function of the electron energy, temperature etc. with a study of the conditions required for the possibility of definition of a general time of relaxation.

The first part of the problem has been treated by Wilson (1936) who considered a general band shape. Putting ϵ a function only of $|\mathbf{k}|$, one finds that the case of the spherically symmetric band is obtained from that of the standard band by the replacement of m^* by $\hbar^2k(d\epsilon/dk)^{-1}$ (the magnitude of the crystal momentum divided by the magnitude of the velocity). Thus, for example, the formulae for the conductivity σ, the Hall constant R and the thermoelectric power Q are

$$\left.\begin{aligned} \sigma &= \frac{e^2}{3\pi^2\hbar^2}\mathscr{K}_1 \\ R &= \frac{3\pi^2}{ec}\frac{1}{\mathscr{K}_1{}^2}\int \tau^2 k\left(\frac{d\epsilon}{dk}\right)^2\frac{\partial f_0}{\partial\epsilon}\,d\epsilon \\ Q &= -\frac{k_0}{e}\left\{\frac{1}{k_0 T}\frac{\mathscr{K}_2}{\mathscr{K}_1}-\frac{\zeta}{k_0 T}\right\} \\ \text{with}\qquad \mathscr{K}_n &= -\int \tau k^2\frac{d\epsilon}{dk}\epsilon^{n-1}\frac{\partial f_0}{\partial\epsilon}\,d\epsilon. \end{aligned}\right\} \quad \ldots\ldots(1)$$

Here, k_0 is Boltzmann's constant and ζ is the Fermi level measured from the bottom of the band. The other symbols have their usual meaning.

The second part of the problem, the determination of τ, is not so straightforward. If, however, the scattering process is elastic we can show quite easily that again m^* for the standard band case will be replaced by $\hbar^2 k(d\epsilon/dk)^{-1}$ for the non-standard band. Thus the usual effective mass tensor is

$$\left(\frac{1}{m^*}\right)_{\mu\nu} = \frac{1}{\hbar^2}\frac{1}{k}\frac{d\epsilon}{dk}\delta_{\mu\nu} + \frac{1}{\hbar^2 k^2}\left(\frac{d^2\epsilon}{dk^2}-\frac{1}{k}\frac{d\epsilon}{dk}\right)k_\mu k_\nu. \quad \ldots\ldots(2)$$

Then

$$\frac{d\mathbf{v}}{dt} = \frac{1}{\hbar^2 k}\frac{d\epsilon}{dk}\mathbf{F} + \frac{1}{\hbar^2 k^2}\left(\frac{d^2\epsilon}{dk^2}-\frac{1}{k}\frac{d\epsilon}{dk}\right)(\mathbf{F}\,.\,\mathbf{k})\mathbf{k}$$

where $\mathbf{F}$ is the force causing the transition, $\mathbf{v}$ is the velocity and t the time. Since the scattering is elastic and the surfaces of constant energy are spheres, $\mathbf{F}$ is tangential to the sphere (being in the direction of $d\mathbf{k}/dt$) and $\mathbf{k}$ is normal to the sphere so that $(\mathbf{F}\,.\,\mathbf{k})=0$ and

$$\frac{d\mathbf{v}}{dt} = \frac{1}{\hbar^2}\frac{1}{k}\frac{d\epsilon}{dk}\mathbf{F}.$$

Therefore in the expression for τ we replace m^* in the standard band case by $\hbar^2 k(d\epsilon/dk)^{-1}$ and this will be equivalent to the use of the usual effective mass tensor.

For inelastic scattering processes, it is not clear what replacement should be made and each case has to be considered separately.

In this paper we shall work out the expression for τ in the case of a spherically symmetric band for two types of scattering processes. First, as an example of an elastic process, we consider scattering by ionized impurities. (This will enable us to find the range of validity of the foregoing argument for a particular process.) Secondly, as an example of an inelastic process, we take scattering by the non-acoustical modes of the lattice vibrations in a homopolar crystal.

§ 3. SCATTERING BY IONIZED IMPURITIES

We approach this problem by the method described by Koster (1954). If we assume that the potential is slowly varying, the coefficients of the Wannier functions in the expansion of the perturbed wave function are obtained from the continuous function $U(\mathbf{r})$ which satisfies the equation

$$\epsilon(-i\nabla)U(\mathbf{r}) + V(\mathbf{r})U(\mathbf{r}) = EU(\mathbf{r}) \quad \ldots\ldots(3)$$

where $V(\mathbf{r})$ is the perturbing potential due to an ionized impurity, E is the perturbed energy and $\epsilon(-i\nabla)$ is $\epsilon(\mathbf{k})$ with $\mathbf{k}$ replaced by $-i\nabla$. The assumptions made here are the same as those in the case of the standard band so that the final results can be compared.

The solutions of the unperturbed problem are the plane waves $U(\mathbf{r})=\exp(i\mathbf{k}\,.\,\mathbf{r})$ with corresponding energy eigenvalues $\epsilon(\mathbf{k})$. We want an asymptotic solution of (3) of the form

$$U(\mathbf{r})=\exp(i\mathbf{k}\,.\,\mathbf{r})+\frac{\exp(ikr)}{r}f(\theta,\phi) \qquad \ldots\ldots(4)$$

with $E=\epsilon(k)$.

Let us try a solution of the form

$$U(\mathbf{r})=\exp(i\mathbf{k}\,.\,\mathbf{r})+v(\mathbf{r}). \qquad \ldots\ldots(5)$$

The equation for $v(\mathbf{r})$ is

$$\{\epsilon(-i\nabla)-\epsilon(\mathbf{k})\}v(\mathbf{r})=-V(\mathbf{r})\exp(i\mathbf{k}\,.\,\mathbf{r}). \qquad \ldots\ldots(6)$$

The Green's function for equation (6) is

$$\begin{aligned} G(\mathbf{r},\mathbf{r}') &= \frac{1}{(2\pi)^3}\int\frac{\exp\{i\mathbf{k}\,.\,(\mathbf{r}-\mathbf{r}')\}}{\epsilon(\mathbf{k}')-\epsilon(\mathbf{k})}\,d\mathbf{k}' \\ &= \frac{1}{2\pi^2\,|\mathbf{r}-\mathbf{r}'|}\int_0^\infty\frac{k'dk'\sin(k'|\mathbf{r}-\mathbf{r}'|)}{\epsilon(k')-\epsilon(k)}. \end{aligned} \qquad \ldots\ldots(7)$$

Let us assume that the integrand is analytic at all points on the real axis of k'-space except for simple poles at $k'=\pm k$. This means that the equation in k, $\epsilon(k)=$ constant, has only two real simple solutions. With this assumption, we shall evaluate the integral in (7) and shall, for the moment, ignore the other poles, and possible branch points, of the integrand at complex values of k'. We shall return to this subject at the end of the section.

The integral in (7) can be evaluated in the usual way, choosing the contour of integration in such a way as to give only outgoing waves. We find

$$G(\mathbf{r},\mathbf{r}')=\frac{\exp(ik\,|\mathbf{r}-\mathbf{r}'|}{2\pi\,|\mathbf{r}-\mathbf{r}'|}\,k\left(\frac{d\epsilon}{dk}\right)^{-1}. \qquad \ldots\ldots(8)$$

Therefore

$$U(\mathbf{r})\xrightarrow[r\to\infty]{}\exp(i\mathbf{k}\,.\,\mathbf{r})-\frac{Ik}{2\pi}\left(\frac{d\epsilon}{dk}\right)^{-1}\frac{\exp(ikr)}{r} \qquad \ldots\ldots(9)$$

where

$$I=\int\exp(-ikr'\cos\theta)\,V(\mathbf{r}')\exp(i\mathbf{k}\,.\,\mathbf{r}')\,d\mathbf{r}'$$

and θ is the angle of scattering.

If we choose $V(\mathbf{r})=e^2/\kappa r)\exp(-qr)$ where κ is the dielectric constant of the material and q is a screening constant, then

$$I=-\frac{4\pi e^2}{\kappa}\,\frac{1}{q^2+4k^2\sin^2\tfrac{1}{2}\theta}$$

and the cross section for scattering through an angle θ is

$$\sigma(\theta)=\frac{e^4}{\kappa^2}\,4k^2\left(\frac{d\epsilon}{dk}\right)^{-2}(q^2+4k^2\sin^2\tfrac{1}{2}\theta)^{-2}. \qquad \ldots\ldots(10)$$

From this, τ is obtained in the usual way and we find

$$\frac{1}{\tau}=\frac{2\pi}{\hbar}\frac{e^4}{\kappa^2}N\frac{1}{k^2d\epsilon/dk}\left[\ln\left(1+\frac{4k^2}{q^2}\right)-\frac{4k^2}{q^2+4k^2}\right]. \qquad \ldots\ldots(11)$$

Here N is the concentration of scattering centres.

For the standard band, Mott (1936) found

$$\frac{1}{\tau}=\frac{2\pi}{\hbar}\frac{e^4}{\kappa^2}N\frac{m^*}{\hbar^2k^3}\left[\ln\left(1+\frac{4k^2}{q^2}\right)-\frac{4k^2}{q^2+4k^2}\right].$$

Thus the non-standard band formula is obtained from the standard band one by expressing the latter in terms of k and m^*, then replacing m^* by $\hbar^2k(d\epsilon/dk)^{-1}$.

Let us now return to the assumption made about the form of $\epsilon(k)$. The following cases may arise.

(i) The equation $\epsilon(k)=\text{constant}$ may have more than two real solutions for k. It would then describe a many-valley band structure and there would be contributions to the integral in (7) from the other poles in the real axis of k'-space. These would give additional terms in equations (8) and (9) describing the type of inter-valley scattering discussed by Herring (1955) for germanium. We have chosen to neglect this type of scattering and have chosen only forms of $\epsilon(k)$ which are monotonically increasing with k.

(ii) If $\epsilon(k)$ is monotonically increasing, and not of the form $\epsilon=\hbar^2k^2/2m^*$, there will be other solutions of the equation $\epsilon(k)=\text{constant}$ for complex values of k. Thus the integral in (7) will have poles at complex values of k' and these will lead to additional terms in (8) and (9) which decay exponentially as r tends to infinity. (These are the non-propagating solutions discussed by Koster (1954).) For scattering by a single impurity centre these terms can then be ignored. If, however, the concentration of impurity centres is sufficiently great, these terms will not have decayed to zero by the time the scattered wave reaches the next impurity centre. As an example, if $\epsilon=\alpha k^4$, these terms can be ignored when the value of k considered is less than both q and r_0^{-1}, where q is the screening constant of the potential and r_0 is the distance between the impurities.

(iii) The function $\epsilon(k)$ may also be such that the integrand in (7) has branch points at some complex values of k'. These will again lead to exponentially damped scattered waves. As an example, if $\epsilon=(k^2+a^2)^{1/2}-a$, the terms arising from the branch points at $k'=\pm ia$ can be ignored if a is less than both q and r_0^{-1}. If the integrand has branch points on the real axis of k'-space, no general rule can be given for the treatment of the integral in (7) and each of these cases would have to be treated independently.

(iv) The other possibility for the form of $\epsilon(k)$ is that it is such that the poles at $k'=\pm k$ are poles of higher order than the first. If this occurs, then in equation (8), we would have higher derivatives of ϵ with respect to k.

We have thus given a full quantum-mechanical proof, for the case of scattering by ionized impurities, of the statement made in § 2 that m^* for the standard band should be replaced by $\hbar^2k(d\epsilon/dk)^{-1}$ for elastic scattering processes. It is true, however, only if $\epsilon(k)$ is a monotonic function of k with no branch points for real values of k and if the equation $\epsilon(k)=\text{constant}$ has no real multiple roots. A similar proof of this statement has been given by Radcliffe (1955) for scattering by the acoustical modes of the lattice vibrations, the scattering being assumed to be elastic. The restrictions on the form of $\epsilon(k)$ are inherent in his analysis also.

§ 4. SCATTERING BY NON-ACOUSTICAL MODES

As an example of an inelastic process we consider scattering by the non-acoustical modes of the lattice vibrations in a homopolar crystal. We shall base our analysis on that of Seitz (1948), but will perform the analysis from first principles so that we can see exactly under what conditions a unique time of relaxation exists; this aspect of the problem was not considered by Seitz. The case of scattering by the optical modes in a polar lattice can be dealt with similarly, merely generalizing the work of Howarth and Sondheimer (1953).

Let $P(\mathbf{k}+\mathbf{q}, \mathbf{k})$ and $P(\mathbf{k}-\mathbf{q}, \mathbf{k})$ be the probabilities per unit time of an electron being scattered from the state $\mathbf{k}$ to the states $\mathbf{k}+\mathbf{q}$ and $\mathbf{k}-\mathbf{q}$ respectively. Here $\mathbf{q}$ is the wave-number vector of a phonon.

Let
$$f(\mathbf{k}) = f_0(\mathbf{k}) - \Phi(\mathbf{k}) \frac{\partial f_0}{\partial \epsilon} \qquad \ldots\ldots(12)$$
where f is the distribution function of the electrons and f_0 is its equilibrium value, the usual Fermi function.

Then the rate of change of f due to collisions is
$$\begin{aligned}\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \frac{V}{(2\pi)^3}\frac{1}{k_0 T}\int d\mathbf{q}\,[&P(\mathbf{k}+\mathbf{q}, \mathbf{k}) f_0(\mathbf{k})\{1-f_0(\mathbf{k}+\mathbf{q})\}\{\Phi(\mathbf{k}+\mathbf{q})-\Phi(\mathbf{k})\} \\ &+ P(\mathbf{k}-\mathbf{q}, \mathbf{k}) f_0(\mathbf{k})\{1-f_0(\mathbf{k}-\mathbf{q})\}\{\Phi(\mathbf{k}-\mathbf{q})-\Phi(\mathbf{k})\}]. \qquad \ldots\ldots(13)\end{aligned}$$
Here V is the volume of the crystal and T is the absolute temperature.

The transition probabilities per unit time for absorption and emission of a phonon will be of the form
$$\left.\begin{aligned} P(\mathbf{k}+\mathbf{q}, \mathbf{k}) &= \frac{\pi I^2}{NM\omega} n(\omega)\delta\{\epsilon(|\mathbf{k}+\mathbf{q}|)-\epsilon(k)-\hbar\omega\} \\ P(\mathbf{k}-\mathbf{q}, \mathbf{k}) &= \frac{\pi I^2}{NM\omega}\{n(\omega)+1\}\delta\{\epsilon(|\mathbf{k}-\mathbf{q}|)-\epsilon(k)+\hbar\omega\}. \end{aligned}\right\} \qquad \ldots\ldots(14)$$

In these, N is the number of unit cells in the crystal, M is the mass of the atoms of the crystal, $\hbar\omega$ is the energy of the phonon absorbed or emitted in the transition and $n(\omega)$ is the number of phonons in the mode ω. We take ω to be independent of $\mathbf{q}$. The factor I depends on the matrix element of the collision operator between the initial and final electron wave functions, these wave functions not being reliably known. Seitz (1948) assumes that I is generally a constant parameter, D' say, independent of $\mathbf{q}$. In some cases, however, the symmetry of the electron wave function is such that we have to assume that I depends linearly on $|\mathbf{q}|$ and Seitz takes $I = \frac{2}{3}C'|\mathbf{q}|$ for these cases.†
We therefore consider two cases: case I, $I = D'$; case II, $I = \frac{2}{3}C'|\mathbf{q}|$.

Case I.

Changing $\mathbf{q}$ to $-\mathbf{q}$ in some of the integrals we find from (13)
$$\begin{aligned}\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \frac{V}{(2\pi)^3}\frac{\pi D'^2}{NM\omega}\frac{1}{k_0 T}\int d\mathbf{q}\, &f_0(\mathbf{k})\{1-f_0(\mathbf{k}+\mathbf{q})\}\{\Phi(\mathbf{k}+\mathbf{q})-\Phi(\mathbf{k})\} \\ &\times [n(\omega)\delta\{\epsilon(|\mathbf{k}+\mathbf{q}|)-\epsilon(k)-\hbar\omega\} + \{n(\omega)+1\}\delta\{\epsilon(|\mathbf{k}+\mathbf{q}|)-\epsilon(k)+\hbar\omega\}]. \\ &\qquad \ldots\ldots(15)\end{aligned}$$

Let
$$\Phi(\mathbf{k}) = \frac{\partial \epsilon}{\partial k_x} c\{\epsilon(k)\} = \frac{k_x}{k}\frac{d\epsilon}{dk} c\{\epsilon(k)\}. \qquad \ldots\ldots(16)$$

---

† For further details of the properties of this matrix element, the reader is referred to Seitz's paper.

Let $\mathbf{k}$ be the polar axis in $\mathbf{q}$-space, so that $\mathbf{q}=(q,\theta,\phi)$. Then

$$q_x = q(\cos\theta_k \cos\theta + \sin\theta_k \sin\theta\cos\phi)$$

where θ_k is the angle between $\mathbf{k}$ and the x axis. The integration over ϕ is the first to be performed and for this we need only

$$\int_0^{2\pi} d\phi = 2\pi \text{ and } \int_0^{2\pi} \cos\phi \, d\phi = 0.$$

This integration gives

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \frac{V}{4\pi}\,\frac{D'^2}{NM\omega}\,\frac{k_x}{k_0T}\int q^2\,dq\sin\theta\,d\theta\, f_0(\mathbf{k})\{1-f_0(\mathbf{k}+\mathbf{q})\}$$
$$\times\left[\frac{k+q\cos\theta}{k|\mathbf{k}+\mathbf{q}|}\left(\frac{d\epsilon}{dk}\right)_{k=|\mathbf{k}+\mathbf{q}|} c\{\epsilon(|\mathbf{k}+\mathbf{q}|)\} - \frac{1}{k}\frac{d\epsilon}{dk}c\{\epsilon(k)\}\right]$$
$$\times[n(\omega)\delta\{\epsilon(|\mathbf{k}+\mathbf{q}|)-\epsilon(k)-\hbar w\} + \{n(\omega)+1\}\delta\{\epsilon(|\mathbf{k}+\mathbf{q}|)-\epsilon(k)+\hbar\omega\}]. \quad \ldots\ldots(17)$$

Now
$$\frac{d\{\epsilon(|\mathbf{k}+\mathbf{q}|)-\epsilon(k)\mp\hbar\omega\}}{d(\cos\theta)} = \left(\frac{d\epsilon}{dk}\right)_{k=|\mathbf{k}+\mathbf{q}|}\frac{kq}{|\mathbf{k}+\mathbf{q}|}. \quad \ldots\ldots(18)$$

This enables us to carry out the integration over θ to obtain

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \frac{V}{4\pi}\,\frac{D'^2}{NM\omega}\,\frac{1}{k_0T}\,\frac{k_x}{k}\int_{q_1}^{q_2} q\,dq\left[f_0(\epsilon)\{1-f_0(\epsilon+\hbar\omega)\}\right.$$
$$\times\left\{\frac{k+q\cos\theta_{0,1}}{k}c(\epsilon+\hbar\omega) - \frac{|\mathbf{k}+\mathbf{q}|_{0,1}}{(d\epsilon/dk)_{k=|\mathbf{k}+\mathbf{q}|_{0,1}}}\frac{1}{k}\frac{d\epsilon}{dk}c(\epsilon)\right\}n(\omega)$$
$$+f_0(\epsilon)\{1-f_0(\epsilon-\hbar\omega)\}\left\{\frac{k+q\cos\theta_{0,2}}{k}c(\epsilon-\hbar\omega)\right.$$
$$\left.\left.-\frac{|\mathbf{k}+\mathbf{q}|_{0,2}}{(d\epsilon/dk)_{k=|\mathbf{k}+\mathbf{q}|_{0,2}}}\frac{1}{k}\frac{d\epsilon}{dk}c(\epsilon)\right\}\{n(\omega)+1\}\right] \quad \ldots\ldots(19)$$

where $\theta_{0,1}$ and $\theta_{0,2}$ are such that

$$\left.\begin{aligned}\epsilon(k^2+q^2+2kq\cos\theta_{0,1})-\epsilon(k^2)-\hbar\omega&\equiv 0\\ \epsilon(k^2+q^2+2kq\cos\theta_{0,2})-\epsilon(k^2)+\hbar\omega&\equiv 0\end{aligned}\right\} \quad \ldots\ldots(20)$$

and
$$|\mathbf{k}+\mathbf{q}|_{0,i}{}^2 = k^2+q^2+2kq\cos\theta_{0,i}, \qquad i=1,2.$$

Assuming the same form for $\epsilon(k)$ as in § 3, each of the equations (20) gives only one value of θ_0. From equations (20) we see that $q^2+2kq\cos\theta_{0,i}$ is independent of q and can be treated as a constant in the next integration, that over q.

Let us put
$$k^2+q^2+2kq\cos\theta_{0,i} = k_{f,i} \quad \ldots\ldots(21)$$

so that $\epsilon(k_{f,1})=\epsilon+\hbar\omega$, $\epsilon(k_{f,2})=\epsilon-\hbar\omega$ i.e. $k_{f,1}$ is the radius of the sphere of constant energy $\epsilon+\hbar\omega$ and $k_{f,2}$ is the radius of the sphere of constant energy $\epsilon-\hbar\omega$. Then

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \frac{V}{4\pi}\,\frac{D'^2}{NM\omega}\,\frac{1}{k_0T}\,\frac{k_x}{k}\int_{q_1}^{q_2} q\,dq\left[f_0(\epsilon)\{1-f_0(\epsilon+\hbar\omega)\}\right.$$
$$\times\left\{\frac{(k^2+k_{f,1}{}^2-q^2)}{2k^2}c(\epsilon+\hbar\omega) - \frac{k_{f,1}}{(d\epsilon/dk)_{k=k_{f,1}}}\frac{1}{k}\frac{d\epsilon}{dk}c(\epsilon)\right\}n(\omega)$$
$$+f_0(\epsilon)\{1-f_0(\epsilon-\hbar\omega)\}\left\{\frac{(k^2+k_{f,2}{}^2-q^2)}{2k^2}c(\epsilon-\hbar\omega)\right.$$
$$\left.\left.-\frac{k_{f,2}}{(d\epsilon/dk)_{k=k_{f,2}}}\frac{1}{k}\frac{d\epsilon}{dk}c(\epsilon)\right\}\{n(\omega)+1\}\right]. \quad \ldots\ldots(22)$$

The limits for q are obtained from the fact that $|\cos\theta_{0,i}|\leqslant 1$. The limits for the first set of integrals ($\propto n(\omega)$) are found to be $q_1=k_{f,1}-k$, $q_2=k_{f,1}+k$.

The limits for the second set of integrals ($\propto n(\omega)+1$) are found to be

$$q_1=k-k_{f,2},\ q_2=k+k_{f,2}$$

provided $\epsilon\geqslant\hbar\omega$. If $\epsilon<\hbar\omega$ the second integral gives no contribution. The integration of q is simple and gives

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}}=-\frac{V}{2\pi}\frac{D'^2}{NM\omega}\frac{1}{k_0T}\frac{k_x}{k}\frac{d\epsilon}{dk}c(\epsilon)f_0(\epsilon)\left[\{1-f_0(\epsilon+\hbar\omega)\}n(\omega)\frac{k_{f,1}{}^2}{(d\epsilon/dk)_{k=k_{f,1}}}\right.$$
$$\left.+h(\epsilon-\hbar\omega)\{1-f_0(\epsilon-\hbar\omega)\}\{n(\omega)+1\}\frac{k_{f,2}{}^2}{(d\epsilon/dk)_{k=k_{f,2}}}\right]\quad\ldots\ldots(23)$$

where $h(x)$ is Heaviside's unit function, $h(x)=0$ if $x<0$, $h(x)=1$ if $x>0$. If $\hbar\omega\ll k_0T$, $f_0(\epsilon\pm\hbar\omega)$ can be approximated by $f_0(\epsilon)$ and $n(\omega)$ and $n(\omega)+1$ by $k_0T/\hbar\omega$. Then

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}}=-(f-f_0)\frac{D'^2}{2\pi NM\omega}\frac{k_0T}{\hbar\omega}\left[\frac{k_{f,1}{}^2}{(d\epsilon/dk)_{k=k_{f,1}}}+h(\epsilon-\hbar\omega)\frac{k_{f,2}{}^2}{(d\epsilon/dk)_{k=k_{f,2}}}\right]\,..(24)$$

so that a time of relaxation exists and is given by

$$\frac{1}{\tau}=\frac{D'^2V}{2\pi NM\omega}\frac{k_0T}{\hbar\omega}\left[\frac{k_{f,1}{}^2}{(d\epsilon/dk)_{k=k_{f,1}}}+h(\epsilon-\hbar\omega)\frac{k_{f,2}{}^2}{(d\epsilon/dk)_{k=k_{f,2}}}\right].\quad\ldots\ldots(25)$$

For the standard band (Seitz 1948) the same condition is required for the existence of a time of relaxation and

$$\frac{1}{\tau}=\frac{D'^2V}{2\pi NM\omega}\frac{k_0T}{\hbar\omega}\left[\frac{m^*}{\hbar^2}\,k_{f,1}+h(\epsilon-\hbar\omega)\frac{m^*}{\hbar^2}\,k_{f,2}\right].$$

Thus to obtain the non-standard band formula from the standard, one expresses the latter in terms of final state parameters and replaces m^* by $\hbar^2k(d\epsilon/dk)^{-1}$ evaluated on the appropriate sphere.

Case II.

The analysis for the second assumption as to the value of the transition probabilities is similar to the above except that the final integrations over q are different, the integrands being multiplied by q^2. One finds

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}}=-\frac{2}{27\pi}\frac{VC'^2}{NM\omega}\frac{1}{k_0T}\frac{k_x}{k}\left[n(\omega)f_0(\epsilon)\{1-f_0(\epsilon+\hbar\omega)\}k_{f,1}{}^2\right.$$
$$\times\left\{2kk_{f,1}c(\epsilon+\hbar\omega)-\frac{3d\epsilon/dk}{(d\epsilon/dk)_{k=k_{f,1}}}c(\epsilon)(k^2+k_{f,1}{}^2)\right\}$$
$$+h(\epsilon-\hbar\omega)\{n(\omega)+1\}f_0(\epsilon)\{1-f_0(\epsilon-\hbar\omega)\}k_{f,2}{}^2$$
$$\left.\times\left\{2kk_{f,2}c(\epsilon-\hbar\omega)-\frac{3d\epsilon/dk}{(d\epsilon/dk)_{k=k_{f,2}}}c(\epsilon)(k^2+k_{f,2}{}^2)\right\}\right].\quad\ldots\ldots(26)$$

In order to be able to define a time of relaxation, not only must we assume that $\hbar\omega\ll k_0T$ as before, but we must also assume that $\hbar\omega\ll\epsilon$. Then $c(\epsilon\pm\hbar\omega)$ can be expanded in powers of $\hbar\omega/\epsilon$, $h(\epsilon-\hbar\omega)=1$ and we find

$$\frac{1}{\tau}=\frac{2}{27\pi}\frac{C'^2V}{NM\omega}\frac{k_0T}{\hbar\omega}\left[k_{f,1}{}^2\left\{\frac{2k}{d\epsilon/dk}k_{f,1}+\frac{3(k^2+k_{f,1}{}^2)}{(d\epsilon/dk)_{k=k_{f,1}}}\right\}\right.$$
$$\left.+k_{f,2}{}^2\left\{\frac{2k}{d\epsilon/dk}k_{f,2}+\frac{3(k^2+k_{f,2}{}^2)}{(d\epsilon/dk)_{k=k_{f,2}}}\right\}\right].\quad\ldots\ldots(27)$$

In this case, one has to use $\hbar^2 k(d\epsilon/dk)^{-1}$ evaluated on all three spheres of constant energy ϵ, $\epsilon+\hbar\omega$ and $\epsilon-\hbar\omega$.

For the standard band, $\epsilon=\hbar^2k^2/2m^*$, this reduces to the formula given by Seitz, viz.

$$\frac{1}{\tau}=\frac{2}{27\pi}\frac{C'^2V}{NM\omega}\frac{k_0T}{\hbar\omega}\frac{m^*}{\hbar^2}k^3\left[\left(1+\frac{\hbar\omega}{\epsilon}\right)^{1/2}\left(8+\frac{5\hbar\omega}{\epsilon}\right)+\left(1-\frac{\hbar\omega}{\epsilon}\right)^{1/2}\left(8-\frac{5\hbar\omega}{\epsilon}\right)\right].$$

Since, however, one has to assume that $\hbar\omega\ll\epsilon$, in order to be able to define a time of relaxation, one can expand the radicals to get

$$\frac{1}{\tau}=\frac{2}{27\pi}\frac{C'^2V}{NM\omega}\frac{k_0T}{\hbar\omega}\frac{m^*}{\hbar^2}k^3\left\{16+3\left(\frac{\hbar\omega}{\epsilon}\right)^2\right\}.$$

Once $k_{f,1}$ and $k_{f,2}$ have been expressed in terms of ϵ and $\hbar\omega$, equation (27) can be similarly expanded.

§ 5. DISCUSSION

In a conduction band with a minimum at $\mathbf{k}=0$ in the Brillouin zone, the approximation $\epsilon=\hbar^2k^2/2m^*$ becomes invalid for sufficiently large values of k. Experiments on impure samples of InSb (Hrostowski *et al.* 1954, Barrie and Edmond 1955) have indicated that the conduction band does in fact become filled to such an extent that the quadratic approximation is no longer valid. The simplest assumption to make when this approximation no longer holds is that ϵ is still a function only of $|\mathbf{k}|$, but no longer quadratic. This paper has dealt with the theory of transport phenomena for such bands. It has been shown that the parameter m^* occurring in formulae based on the standard band is usually replaced by $\hbar^2k(d\epsilon/dk)^{-1}$ evaluated at an appropriate energy.

The next approximation to the quadratic form is to take a band of the form $\epsilon=\hbar^2k^2(1-\beta k^2)/2m^*$ and we shall now, as an example, calculate the values of various quantities for this band and compare them with the corresponding values calculated for the standard band $\epsilon=\hbar^2k^2/2m^*$.

For this non-standard band, a surface of constant energy in $\mathbf{k}$-space consists of a pair of concentric spheres and to satisfy the requirements on the form of ϵ stipulated in § 3, we shall consider only the sphere of smaller radius, i.e. we shall use

$$k^2=\frac{1}{2\beta}\left[1-\left(1-\frac{8m^*\beta\epsilon}{\hbar^2}\right)^{1/2}\right]$$

and restrict values of k so that $\beta k^2\leqslant\frac{1}{2}$.

To obtain numerical values, let us take $\beta=5\times10^{17}\,\mathrm{cm}^2$, m^* to be the free electron mass and let us assume that we have 10^{21} electrons per cm^3 in the band. We shall use the formulae appropriate to complete Fermi–Dirac degeneracy.

From the relation $n=k_\zeta{}^3/3\pi^2$, valid for both band forms, we find $\zeta_s=0{\cdot}107$ ev and $\rho_{ns}=0{\cdot}092$ ev. Here n is the concentration of electrons, k_ζ is the value of k at the Fermi energy ρ and the subscripts s and ns refer to the standard and non-standard bands respectively. We shall compare the times of relaxation for scattering by ionized impurities, the more important of the scattering processes discussed in the paper. If the screening constant q is calculated by the Thomas–Fermi method, assuming unit dielectric constant, we find $q_s=1{\cdot}12\times10^8\,\mathrm{cm}^{-1}$ and $q_{ns}=1{\cdot}32\times10^8\,\mathrm{cm}^{-1}$. Using these values of q, we find $(\tau_\zeta)_{ns}=1{\cdot}15(\tau_\zeta)_s$. For both band forms, the Hall constant $R=-1/nec$. The ratio of Hall mobilities is $(-R\sigma)_{ns}/(-\mathrm{R}\sigma)_s=0{\cdot}83$, so that although the time of relaxation is greater in the case of the non-standard band, the Hall mobility is less.

ACKNOWLEDGMENTS

The author would like to express his thanks to Drs. G. Rickayzen and E. H. Sondheimer for useful discussions and to the Admiralty for permission to publish this work.

REFERENCES

BARRIE, R., and EDMOND, J. T., 1955, *J. Electronics* **1,** 161.
HERRING, C., 1955, *Bell Syst. Tech. J.*, **34,** 237.
HOWARTH, D. J., and SONDHEIMER, E. H., 1953, *Proc. Roy. Soc.* A, **219,** 53.
HROSTOWSKI, H. J., WHEATLEY, G. H., and FLOOD, W. F., Jr., 1954, *Phys. Rev.*, **95,** 1683.
KOSTER, G. F., 1954, *Phys. Rev.*, **95,** 1436.
MOTT, N. F., 1936, *Proc. Camb. Phil. Soc.*, **32,** 281.
PARMENTER, R. H., 1955, *Phys. Rev.*, **97,** 587.
RADCLIFFE, J. M., 1955, *Proc. Phys. Soc.* A, **68,** 675.
SEITZ, F., 1948, *Phys. Rev.*, **73,** 549.
WILSON, A. H., 1936, *The Theory of Metals*, 1st Edn (Cambridge : University Press).

The Frequency Response of Optical Systems

By H. H. HOPKINS

Department of Physics, Imperial College, London, S.W.7.

MS. received 14*th September* 1955

Abstract. In optical systems the pencils forming images of extra-axial object points are generally oblique to both the object and image planes. The diffraction image of a point source observed in an oblique image plane differs significantly from that obtained in a plane normal to the image-forming pencil. The theory of the frequency response of optical systems is formulated for this more general case. The resulting formulae are used to relate the response obtained between object and image planes perpendicular to the optical axis to that obtained between planes perpendicular to the principal rays. It is shown that no loss of generality is introduced by the restriction to unidimensional objects. Some general properties of the response function for optical systems having aberration are derived.

§ 1. Formulation of the Problem

Optical systems are frequently required to form images of objects of large angular size, and this usually implies that the principal rays of the pencils forming image points in the outer parts of the field have correspondingly large inclinations to the optical axis. Indeed, the inclinations of these rays may exceed 30 degrees. The diffraction image of an object point observed in an image plane perpendicular to the optical axis may then be significantly different from that found in the corresponding plane perpendicular to the principal ray of the image-forming pencil. This fact is usually ignored in calculating light distributions in the images of extended objects (see, for example, Rayleigh 1896, Duffieux 1946, Hopkins 1953); but, as we shall see, the approximation implied is not generally permissible, even for an aberration-free system.

The geometry of the situation is illustrated in figure 1, in which $E'Q_0'Q'$ is a principal ray emerging from an axially symmetrical optical system whose optical axis is $E'O'$. The location of the geometrical image point in the image plane through O' is conventionally taken to be at the point of intersection Q_0' of the principal ray with this plane, although, of course, the image of the object point takes the form of a diffracted distribution of light. It is, however, usually convenient to take the principal ray intersection point Q_0' as the origin of coordinates in the image plane when calculating this distribution of light. In figure 1, Σ is a reference sphere, having its centre at Q_0' and of radius $E'Q_0'$. The point E' is located in the exit pupil of the system. The extreme ray of the pencil in the sagittal section intersects Σ in D', which point is at a distance h' from the principal ray $E'Q_0'$. Let any other ray cut the reference sphere Σ in a point, such as M' having rectangular coordinates (X, Y), these latter being measured perpendicular to the meridian and sagittal sections respectively.

We shall employ reduced coordinates (x, y), defined by

$$x = X/h', \quad y = Y/h' \qquad \ldots\ldots(1)$$

for the pupil, and reduced coordinates

$$u' = \frac{2\pi}{\lambda}(n' \sin \alpha')\xi', \qquad v' = \frac{2\pi}{\lambda}(n' \sin \alpha')\eta' \qquad \ldots\ldots(2)$$

for the region of the image plane in the neighbourhood of Q_0'. The angle α' is that subtended by E′D′ at Q_0' and (ξ', η') are geometrical coordinate distances with origin at Q_0', ξ' being parallel to the x axis, and η' lying in the line $O'Q_0'$.

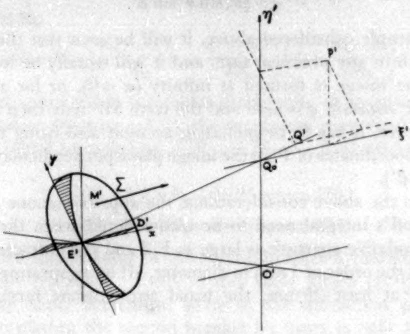


Figure 1.

Let the wave front aberration with reference to the sphere Σ be denoted by $W(x, y)$. The pupil function is then defined by

$$\left.\begin{aligned} f(x,y) &= \tau(x,y)\exp\{ikW(x,y)\} && \text{within A} \\ &= 0 && \text{outside A} \end{aligned}\right\} \qquad \ldots\ldots(3)$$

where A is used to denote the region of the actual pupil, and $k = 2\pi/\lambda$. The factor $\tau(x, y)$ is wholly real, and takes account of any non-uniformity of amplitude over the emergent wave front. It is included here merely for the sake of generality, most of the applications contemplated being those for which we shall put $\tau(x, y) = 1$. In finding the complex amplitude of the diffracted light at a point P′, having coordinates (ξ', η'), we have to note that P′ lies in an image plane perpendicular to the principal ray through Q′, where $Q_0'Q' = \eta' \sin \beta'$, β' being the angle the principal ray makes with the optical axis. Moreover, in this plane, the coordinates of P′ are $(\xi', \eta' \cos \beta')$. The longitudinal focal shift $Q_0'Q'$ introduces an additional aberration term (Hopkins 1950), which is given by

$$\delta W = \tfrac{1}{2} n' \sin^2 \alpha' (\eta' \sin \beta')(x^2 + y^2). \qquad \ldots\ldots(4)$$

The magnitude of this additional term can be appreciably greater than a wavelength. In an extreme case, for example $\beta' = 30°$ and a relative aperture F/2, equation (4) gives $\delta W = \eta'/64$. We shall suppose that, for all systems in question, $R_0 = E'Q_0' \simeq 25$ mm and the aberration is such that the image patch has no significant light intensity outside a circle 1 mm in diameter. This seems to be a reasonable assumption in the light of recently published measurements of energy concentrations in point source images (R. E. Hopkins 1954). The angular size of the image patch is then approximately 2°. Putting $\eta'=0{\cdot}5$ mm and $\lambda=0{\cdot}5\,\mu$ we find $\delta W=15{\cdot}6\,\lambda$, and this term is therefore certainly not negligible. It will be observed, moreover, that the term δW is present even in the absence of aberration in the ordinary sense.

It will be convenient in what follows to write the phase advance corresponding to δW in terms of the reduced coordinate $v'=(2\pi/\lambda)(n'\sin\alpha')\eta'$. We then find, in place of (4),

$$k\delta W=qv'(x^2+y^2) \qquad \ldots\ldots(5)$$

where

$$q=\tfrac{1}{2}n'\sin\alpha'\sin\beta'. \qquad \ldots\ldots(6)$$

From the example considered above, it will be seen that the constant q will never exceed 1/16 in any practical case, and it will usually be less than half this value. When the image is formed at infinity ($\alpha'=0$), or for a telecentric exit pupil ($\beta'=0$), the constant q is zero and the term δW may then be ignored. In most cases, however, it has to be included, account also being taken of the fact that the reduced coordinates of P′ in the image plane perpendicular to the principal ray are $(u', v'\cos\beta')$.

In addition to the above considerations, the approximations usually made in applying Kirchhoff's integral need to be reconsidered when the image-forming wave front has a relative aperture as large as F/2 and the diffracted patch of light is allowed to be of the order of 1 mm in diameter. If we again suppose the distance $E'Q_0'=R_0$ to be at least 25 mm, the usual approximate form of Kirchhoff's integral

$$A(\xi',\eta')=\frac{i}{\lambda}\iint\frac{1+\cos\chi}{2RR_0}\,a(X,Y)\exp\{ik[R_0-R]\}dS \qquad \ldots\ldots(7)$$

is clearly of sufficient accuracy. $a(X,Y)$ is the complex amplitude at the point (X,Y) of the reference sphere (figure 2). R, R_0 are the distances of this point from Q′ and P′ respectively, $k=2\pi/\lambda$, χ is the angle between R and R_0, and dS is an element of area on the reference sphere at the point (X,Y,Z). The square bracket $[R_0-R]$ denotes an optical path length. With the restriction placed on R_0, and $Q'P'\leqslant 0{\cdot}5$ mm, the relative variation in $(1+\cos\chi)/2R_0R$ is small enough to be ignored. The variation in this factor amounts, in any case, to a small non-uniformity of amplitude over the wave front of such magnitude that the effects on the image will be negligible (Hopkins 1949). For a similar reason we may write $dS=dXdY$, and the only significant error remains that implied in the usual expression for the phase factor $k[R_0-R]$. This, we shall find, contains a non-negligible term in $(\xi'^2+\eta'^2)$.

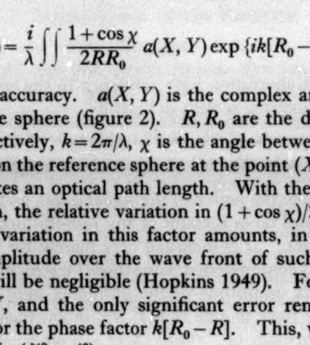


Figure 2.

In figure 2 the origin of the coordinate system (X, Y, Z) is at E′, and P′ is the point (ξ', η', R_0). Hence

$$\begin{aligned} R^2 &= (X-\xi')^2 + (Y-\eta')^2 + (Z-R_0)^2 \\ &= R_0{}^2 + \xi'^2 + \eta'^2 - 2(X\xi' + Y\eta') \end{aligned}$$

the point (X, Y, Z) being on the sphere whose equation is $X^2 + Y^2 + (Z-R_0)^2 = R_0{}^2$. The difference $R_0 - R$ may now be written

$$R_0 - R = \frac{2(X\xi' + Y\eta')}{R_0 + R} - \frac{\xi'^2 + \eta'^2}{R_0 + R}$$

or, since $R_0 - R \ll R_0$,

$$R_0 - R = \left\{\frac{X\xi' + Y\eta'}{R_0} - \frac{\xi'^2 + \eta'^2}{2R_0}\right\}\left(1 + \frac{R_0 - R}{2R_0}\right).$$

If the wave front is limited by a circular diaphragm, of angular radius 0·25 seen from Q′ (i.e. relative aperture = F/2), the maximum numerical value of the first bracket may be found by inserting $X/R_0 = 0{\cdot}25$, $\eta' = 0$, $\xi' = 0{\cdot}5$ mm, $R_0 = 25$ mm, giving

$$(X\xi' + Y\eta')/R_0 = 0{\cdot}125 \text{ mm} \simeq 250\,\lambda$$

$$(\xi'^2 + \eta'^2)/2R_0 = 0{\cdot}005 \text{ mm} \simeq 10\,\lambda.$$

The maximum numerical value of the second bracket is therefore approximately 1·0025, since $R_0 - R \simeq 250\,\lambda$. Even in this case, which corresponds to points in the extremely faint, outermost, parts of the diffraction pattern, the error resulting from replacing the second bracket by unity is only about $0{\cdot}6\,\lambda$. The effect of adding the corresponding phase difference to that associated with $R_0 - R$, which is of the order of $500\,\pi$, would be negligible. For points nearer the centre of the diffraction image this error, which is a smooth function of (X, Y), would be negligible. On the other hand the term $(\xi'^2 + \eta'^2)/2R_0$, which is usually neglected as small, is seen to be significant under conditions occurring in practical systems. In terms of reduced coordinates

$$k[R_0 - R] = u'x + v'y - d(u'^2 + v'^2) \qquad \ldots\ldots(8)$$

in which

$$d = (2kh'n' \sin\alpha')^{-1}.$$

For an off-axis image point such as Q′ the complex amplitude on the reference sphere $a(X, Y)$ is expressed by the pupil function

$$a(X, Y) = f(x, y)\exp(ik\delta W)$$

$f(x, y)$ being defined by (3), and δW by (5) and (6). The reduced coordinates of P′ (figure 1) measured in the plane perpendicular to the principal ray are $(u', v'\cos\beta')$. Using the expressions (5) and (8), the complex amplitude at P′ is therefore given, according to (7), by the expression

$$\begin{aligned} F_0(u', v') = \frac{1}{2\pi}\exp\{-id(u'^2 + v'^2\cos^2\beta')\}\iint_{-\infty}^{+\infty} f(x, y)\exp\{iqv'(x^2 + y^2)\} \\ \times \exp\{i(u'x + v'y\cos\beta')\}dxdy \qquad \ldots\ldots(9) \end{aligned}$$

from which a constant factor is omitted. The factor $1/2\pi$ is inserted for convenience in subsequent calculations. The expression (9) is usually written in the form resulting when $d = 0$, $q = 0$ and $\cos\beta' = 1$. As we have seen, a gross approximation is thereby implied in many cases of practical interest.

We shall be concerned in what follows exclusively with self-luminous (incoherent) objects. The intensity in the diffraction image of a point source in the neighbourhood of P′ will be denoted by

$$G(u',v') = |F_0(u',v')|^2 \qquad \ldots\ldots(10)$$

and is, of course, independent of the exponential factor outside the integral (9). For our purposes, therefore, we may omit this factor and re-define $F_0(u',v')$ by the expression

$$F_0(u',v') = \frac{1}{2\pi}\iint_{-\infty}^{+\infty} f(x,y)\exp\{iqv'(x^2+y^2)\}$$
$$\times \exp\{i(u'x+v'y\cos\beta')\}dxdy \qquad \ldots\ldots(11)$$

although in other cases, namely those of coherent or partially coherent objects, account has to be taken of this factor in integrating the disturbances in the image plane arising from different points of the object. In our case, however, only those additional terms remain which are associated with the obliquity of the image plane to the principal ray.

It is important to note that the correction terms introduced above will be, for practical purposes, constant for points of the image within a radius of 0·5 mm centred on Q_0'. The aberration of the system will also not change appreciably over this region. Since points at distances from Q_0' greater than 0·5 mm contribute nothing to the intensity at Q_0', we may conveniently regard (11) as formally valid for all points of the image plane when calculating the images of extended objects formed in the neighbourhood of Q_0'. The disturbance at P′ due to an element of the object imaged geometrically at a point having coordinates (u, v) with respect to Q_0' will then be given by $G(u'-u, v'-v)$. This fact permits the application of convolution theorems in subsequent calculations relating to extended objects. We shall, however, first need to show that a region of the object at the point conjugate to Q_0' is imaged isometrically in the neighbourhood Q_0', and find the appropriate magnification law.

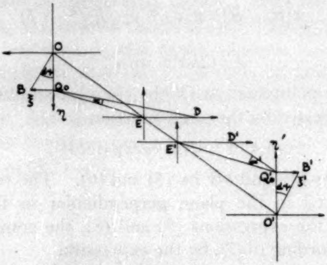


Figure 3.

In figure 3 the small segment Q_0B is in the sagittal section of the principal ray Q_0E. By symmetry, BQ_0 being practically an arc of a circle of centre O, the geometrical image B′ is such that $Q_0'B'$ subtends an angle $d\psi$ at O′ equal to that subtended by QB at O. If $Q_0B = \xi$, $Q_0'B' = \xi'$, $OQ_0 = \bar{\eta}$ and $O'Q_0' = \bar{\eta}'$, the sagittal magnification for the neighbourhood of Q_0' is

$$M = \xi'/\xi = \bar{\eta}'/\bar{\eta}. \qquad \ldots\ldots(12)$$

The magnification in the meridian section will depend on the geometrical distortion in the image. The full curve in figure 4 shows the type of variation of $\bar{\eta}'$ with $\bar{\eta}$ in the presence of distortion terms depending on the third and fifth powers of the image size. It is tangential at the origin to the line $\bar{\eta}'=M_0\bar{\eta}$, M_0 being the magnification in the absence of distortion. This line is shown as OA_0 in the diagram. A typical point Q_0' is marked on the curve, and the line OA, passing through this point, corresponds to an 'undistorted' magnification $M=\bar{\eta}'/\bar{\eta}$, which is equal to the sagittal magnification (12). Consideration of practical cases shows that distortion errors of 10% with respect to image points of magnification M_0 lead to errors of less than 1% with respect to image points of magnification M for the neighbourhood of Q_0'. In consequence we may assume the magnification in the meridian section for the neighbourhood of Q_0' to be equal to that for the point Q_0' itself, and therefore also equal to the sagittal magnification in the neighbourhood of Q_0'. Hence with sufficient accuracy a region around Q is imaged isometrically around Q_0', and the magnification is equal to the sagittal magnification M.

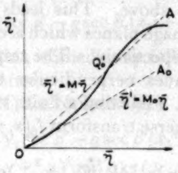


Figure 4.

To find M, we note that the aberration of the image point B′ is the same as that of Q_0'. In consequence the angles subtended by ED, E′D′, at Q, Q_0' respectively, satisfy the sine condition. That is

$$(n'\sin\alpha')\xi'=(n\sin\alpha)\xi \qquad ......(13)$$

and the magnification for the neighbourhood of Q_0' is given by

$$M=\frac{n\sin\alpha}{n'\sin\alpha'}. \qquad ......(14)$$

It follows that, if (u,v) are reduced coordinates defined for the neighbourhood of Q by expressions analogous to (2), the geometrical image of the point (u,v) of the object plane has coordinates $u'=u$, $v'=v$.

Since points at distances from Q_0' greater that 0·5 mm contribute nothing to the intensity at Q_0', we may treat the above magnification relations as valid for the whole object and image planes. Thus, if the intensity at the point (u,v) of the object plane is written $B(u,v)$, the intensity in the image plane is given, apart from a constant factor, by

$$B'(u',v')=\frac{1}{2\pi}\iint_{-\infty}^{+\infty}B(u,v)G(u'-u,v'-v)dudv. \qquad ......(15)$$

$G(u',v')$, the intensity in the point-source diffraction image centred on $u'=v'=0$, is expressed by the formulae (9) and (10) above. The results of calculations using (15) are of course valid only for the neighbourhood of Q_0'. In what follows we shall apply a convolution theorem to (15), and discuss the properties of the inverse transform $g(s,t)$ of $G(u',v')$ which occurs in the resulting product relation.

§ 2. Application of the Convolution Theorem

Following Duffieux (1946) we shall apply the convolution theorem for Fourier transforms to (15). The inverse transforms of the functions occurring there then satisfy the product relation

$$b'(s,t)=g(s,t)b(s,t) \qquad \ldots\ldots(16)$$

the inverse transform $g(s,t)$ being defined by

$$g(s,t)=\frac{1}{2\pi}\iint_{-\infty}^{+\infty} G(u',v')\exp\{-i(u's+v't)\}du'dv' \qquad \ldots\ldots(17)$$

and similarly for $b(s,t)$, $b'(s,t)$. Substituting $|F_0(u',v')|^2$ for $G(u',v')$ in accordance with (10), and applying Parseval's theorem, we obtain

$$g(s,t)=\frac{1}{2\pi}\iint_{-\infty}^{+\infty} f_0(x+\tfrac{1}{2}s, y+\tfrac{1}{2}t)f_0^*(x-\tfrac{1}{2}s, y-\tfrac{1}{2}t)dxdy \qquad \ldots\ldots(18)$$

$f_0(x,y)$ being the inverse transform of $F_0(u',v')$, and the asterisk denoting the complex conjugate. If the customary approximations are made, we put $q=0$ and $\cos\beta'=1$ in (11), and the inverse transform $f_0(x,y)$ is then simply the pupil function $f(x,y)$ defined in (3) above. This leads to a response (transmission) factor defined for object and image planes which are perpendicular to the incident and emergent principal rays respectively. The response factor $g(s,t)$, defined for image formation between planes perpendicular to the optical axis, must be calculated using (18) and (11), admitting a finite value for q and that $\cos\beta'\neq 1$.

From the definition the inverse transform $f_0(x,y)$ of (11) is given by

$$f_0(x,y)=\left(\frac{1}{2\pi}\right)^2\iiiint_{-\infty}^{+\infty} f(x_0,y_0)\exp\{iqv'(x_0^2+y_0^2)\}\exp\{i(u'x_0+v'y_0\cos\beta')\}$$
$$\times\exp\{-i(u'x+v'y)\}du'dv'dx_0dy_0$$

or, grouping terms and integrating with respect to u' and v',

$$f_0(x,y)=\iint_{-\infty}^{+\infty} f(x_0,y_0)\delta\{x_0-x, \cos\beta' y_0+q(x_0^2+y_0^2)-y\}dx_0dy_0 \qquad \ldots\ldots(19)$$

the second factor in the integrand being a delta function. To effect the remaining integrations we replace (x_0,y_0) in $f(x_0,y_0)$ by appropriate roots of the equations

$$x_0-x=0, \qquad \cos\beta' y_0+q(x_0^2+y_0^2)-y=0.$$

Since the value of q never exceeds 1/16, an approximate solution of the second of these equations is $y_0=y/\cos\beta'$. A more accurate solution is obtained by replacing the term qy_0^2 by $qy^2/\cos^2\beta'$, giving

$$x_0=x, \qquad y_0=\frac{y}{\cos\beta'}-\frac{q}{\cos\beta'}\left(x^2+\frac{y^2}{\cos^2\beta'}\right).$$

If the error in this value of y_0 is denoted by ϵ, substitution in the original equation shows that $\epsilon<0{\cdot}01$, even when $q=1/16$. Equation (19) now becomes

$$f_0(x,y)=f(x, y\sec\beta'-q\sec\beta'\{x^2+y^2\sec^2\beta'\}).$$

Returning now to (18), substitution of this expression for $f_0(x,y)$ gives

$$g(s,t)=\frac{1}{2\pi}\iint_{-\infty}^{+\infty} f(x+\tfrac{1}{2}s, \{y+\tfrac{1}{2}t\}\sec\beta'-q\sec\beta'[(x+\tfrac{1}{2}s)^2+(y+\tfrac{1}{2}t)^2\sec^2\beta'])$$
$$\times f^*(x-\tfrac{1}{2}s, \{y-\tfrac{1}{2}t\}\sec\beta'-q\sec\beta'[(x-\tfrac{1}{2}s)^2+(y-\tfrac{1}{2}t)^2\sec^2\beta'])dxdy.$$
$$\ldots\ldots(20)$$

We shall now show that for practical cases, these terms in the square brackets of (20) which involve s, t may be neglected. If, for example, the sagittal ray through D′ (figure 1) is on the periphery of a circular diaphragm whose plane is normal to $E'Q_0'$, the pupil lies in the interior of the circle $x^2+y^2=1$. Hence, by (3), $f(x,y)=f^*(x,y)=0$ for $x^2+y^2>1$. Accordingly, q being small, the maximum value of $(x\pm\frac{1}{2}s)^2+(y\pm\frac{1}{2}t)^2\sec^2\beta'$ is unity. If, therefore, we make the approximation

$$q\sec\beta'[(x\pm\tfrac{1}{2}s)^2+(y\pm\tfrac{1}{2}t)^2\sec^2\beta']=q\sec\beta'[x^2+y^2\sec^2\beta']$$

the error is approximately $\pm q\sec\beta'(xs+yt\sec\beta')$ for small values of s and t, each less than (say) 0·20. This, we shall see, corresponds to detail of 100 lines/mm with an F/2 image-forming cone, and values of s and t exceeding 0·2 are, therefore, of little practical importance at apertures and angles of field corresponding to $q=1/16$. In such systems, therefore, we may employ this approximation, and (20) then becomes

$$g(s,t)=\frac{1}{2\pi}\iint_{-\infty}^{+\infty} f(x_1+\tfrac{1}{2}s, [y_1+\tfrac{1}{2}t]\sec\beta'-q\sec\beta'[x_1^2+y_1^2\sec^2\beta'])$$
$$\times f^*(x_1-\tfrac{1}{2}s, [y_1-\tfrac{1}{2}t]\sec\beta'-q\sec\beta'[x_1^2+y_1^2\sec^2\beta'])dx_1dy_1 \quad \ldots(21)$$

in which, for convenience in what immediately follows, the subscript 1 has been given to the variables of integration.

If the variables (x_1, y_1) in (21) are replaced by

$$x=x_1, \qquad y=y_1\sec\beta'-q\sec\beta'(x_1^2+y_1^2\sec^2\beta') \quad \ldots\ldots(22)$$

the element of area is given by

$$dx_1dy_1=\frac{\partial(x_1,y_1)}{\partial(x,y)}\,dxdy. \quad \ldots\ldots(23)$$

Recalling that q never exceeds 1/16, we may invert (22), obtaining to a good approximation,

$$x_1=x, \qquad y_1=y\cos\beta'+q(x^2+y^2).$$

Evaluation of the Jacobian in (23), and substitution from (22) and (23) in (21), then gives

$$g(s,t)=\frac{\cos\beta'}{2\pi}\iint_{-\infty}^{+\infty} f(x+\tfrac{1}{2}s, y+\tfrac{1}{2}t\sec\beta')f^*(x-\tfrac{1}{2}s, y-\tfrac{1}{2}t\sec\beta')dxdy$$
$$+\frac{q\cos\beta'}{\pi}\iint_{-\infty}^{+\infty} f(x+\tfrac{1}{2}s, y+\tfrac{1}{2}t\sec\beta')f^*(x-\tfrac{1}{2}s, y-\tfrac{1}{2}t\sec\beta')y\,dxdy.$$
$$\ldots\ldots(24)$$

It is convenient to normalize (24) in such a manner that a uniform object with intensity $B(u,v)=1$ gives rise to a uniform image with intensity $B'(u',v')=1$. For this reason we shall replace $g(s,t)$ by

$$D(s,t)=g(s,t)/g(0,0) \quad \ldots\ldots(25)$$

as in previous work (Hopkins 1953, 1955).

The value of $g(0,0)$ is given by

$$g(0,0)=\frac{\cos\beta'}{2\pi}\iint_{-\infty}^{+\infty}|f(x,y)|^2\,dxdy+\frac{q\cos\beta'}{\pi}\iint_{-\infty}^{+\infty}|f(x,y)|^2y\,dxdy$$

in which, if the form and transparency of the pupil $\tau(x,y)$ are symmetrical about the x axis, the second integral is zero. We thus obtain for (25) the form

$$D(s,t)=\frac{1}{A}\iint_{-\infty}^{+\infty} f(x+\tfrac{1}{2}s, y+\tfrac{1}{2}t\sec\beta')f^*(x-\tfrac{1}{2}s, y-\tfrac{1}{2}t\sec\beta')dxdy$$
$$+\frac{2q}{A}\iint_{-\infty}^{+\infty} f(x+\tfrac{1}{2}s, y+\tfrac{1}{2}t\sec\beta')f^*(x-\tfrac{1}{2}s, y-\tfrac{1}{2}t\sec\beta')y\,dx\,dy \quad \ldots\ldots(26)$$

in which

$$A=\iint_{-\infty}^{+\infty} |f(x,y)|^2\,dx\,dy.$$

If the pupil is of constant transparency, we may further write

$$|f(x,y)|^2=\{\tau(x,y)\}^2=1,$$

and A then simply denotes the area of the pupil.

$D(s,t)$ is the normalized response of the optical system to the spatial frequency-pair (s,t) and the normalized form of (16) becomes

$$b'(s,t)=D(s,t)b(s,t), \quad \ldots\ldots(27)$$

inversion of which gives

$$B'(u',v')=\frac{1}{2\pi}\iint_{-\infty}^{+\infty} D(s,t)b(s,t)\exp\{i(u's+v't)\}\,dsdt \quad \ldots\ldots(28)$$

for the intensity distribution in the image of the object $B(u,v)$. From this we see that $|D(s,t)|$ is the ratio of the 'contrast' of the frequency pair (s,t) as found in the image to that of the same frequency pair in the object, and $\arg(D(s,t))$ denotes a spatial phase-shift of this Fourier component relative to the geometrical image.

The significance of the Fourier frequency variables (s,t) may be seen as follows. In one period of the frequency s, the product $u's$ increases by 2π. Thus, recalling the definition (2) of u', we find

$$\left\{\frac{2\pi}{\lambda}(n'\sin\alpha')\xi_0'\right\}s=2\pi$$

ξ_0' being the length of period as it appears in the image. Denoting the reciprocal $1/\xi_0'$ by R' we now find

$$S=\frac{\lambda}{n'\sin\alpha'}R'=\frac{\lambda}{n\sin\alpha}R \quad \ldots\ldots(29)$$

R being similarly defined for the object plane. If λ is in (say) millimetres, R, R' denote the number of lines per millimetre, corresponding to the component of frequency s, as it appears in the object and image planes respectively. A similar significance attaches to the variable t.

If the approximations $q=0$, $\cos\beta'=1$ are made, (26) reverts to the form

$$D(s_0,t_0)=\frac{1}{A}\iint_{-\infty}^{+\infty} f_0(x_0+\tfrac{1}{2}s, y_0+\tfrac{1}{2}t)f_0^*(x_0-\tfrac{1}{2}s, y_0-\tfrac{1}{2}t)\,dx_0dy_0 \quad \ldots\ldots(30)$$

in which the zero subscripts are introduced for convenience in what follows. In this case, when the object and image planes are perpendicular to the principal ray, it has been shown (Hopkins 1955) that the response $D(s_0,t_0)$ for the frequency pair (s_0,t_0) of a two-dimensional object is identical with the response $D(s,\psi)$ for the frequency $s=\sqrt{(s_0^2+t_0^2)}$ of a unidimensional object whose direction is inclined at an angle $\psi=\tan^{-1}(t_0/s_0)$ to the direction of the v axis. This interpretation is useful in practical work, in that it shows that all the image-forming properties of an optical system may be studied experimentally using unidimensional test-objects and scanning apertures in different azimuths.

In analytical work it is convenient to denote the pupil function by $f_0(x_0, y_0)$ and the frequency variables by (s_0, t_0) as in (30), and to evaluate $D(s_0, t_0)$ by calculating the response to the equivalent frequency, namely

$$D(s,\psi)=\frac{1}{A}\iint_{-\infty}^{+\infty} f(x+\tfrac{1}{2}s, y)f^*(x-\tfrac{1}{2}s, y)dxdy \qquad \ldots\ldots(31)$$

in which $f(x, y)$ derives from $f_0(x_0, y_0)$ by the substitutions

$$\left.\begin{aligned} x_0 &= x\cos\psi - y\sin\psi \\ y_0 &= y\cos\psi + x\sin\psi \end{aligned}\right\} \qquad \ldots\ldots(32)$$

which imply a rotation of the pupil axes through an angle $\psi=\tan^{-1}(t_0/s_0)$. The equivalent frequency is given by $s=\sqrt{(s_0^2+t_0^2)}$. The structure implied in (31) is unidimensional, of frequency s, and parallel with the v axes of new coordinates (u, v), (u', v') for the object and image planes respectively. These new axes make an angle ψ with the original axes (u_0, v_0), (u_0', v_0') respectively. To specify the performance of the optical system at any given part of the image plane, it is sufficient to give the response $D(s, \psi)$ as a function of s for different azimuths ψ.

The above result does not arise from the particular analytical expression for the response $D(s_0, t_0)$; it is a consequence of the equivalence of the component $b_0(s_0, t_0)\exp\{i(u_0s_0+v_0t_0)\}$ of the Fourier spectrum of a function $B_0(u_0, v_0)$ of two variables and the component $b(s)\exp\{ius\}$ of the spectrum of a function of the single variable $u=u_0\cos\psi+v_0\sin\psi$, where $\psi=\tan^{-1}(t_0/s_0)$ and $s=\sqrt{(s_0^2+t_0^2)}$. This result is expressed by the identity

$$b_0(s_0, t_0)\exp\{i(u_0s_0+v_0t_0)\}\equiv b_0(s_0, t_0)\exp\{ius\} \qquad \ldots\ldots(33)$$

u and s having the significance given above. Since, by (16), the optical system 'responds' independently to the components of different frequency in the Fourier spectrum of the object intensity function, it will not distinguish between the two-dimensional object giving rise to the left-hand side of (33) and a unidimensional object giving rise to the right-hand side. It follows that the response will be identical in the two cases, and the earlier result $D(s_0, t_0)\equiv D(s, \psi)$ is therefore seen to be valid in all circumstances. In particular, the new factors treated in arriving at the expression (26) will not affect the validity of this identity in either experimental or analytical work.

We may now transform (26), to which are first added appropriate zero suffixes, by means of the substitution (32). The response $D(s_0, t_0)$ is then expressed by the equivalent formula

$$\begin{aligned} D(s,\psi) = {} & \frac{1}{A}\iint_{-\infty}^{+\infty} f(x+\tfrac{1}{2}s, y)f^*(x-\tfrac{1}{2}s, y)dxdy \\ & + \frac{2q}{A}\iint_{-\infty}^{+\infty} f(x+\tfrac{1}{2}s, y)f^*(x-\tfrac{1}{2}s, y)(y\cos\psi - x\sin\psi)dxdy \qquad \ldots(34) \end{aligned}$$

the equivalent frequency being given by

$$s=(s_0^2+t_0^2\sec^2\beta')^{1/2} \qquad \ldots\ldots(35)$$

and the angle ψ is

$$\psi = \tan^{-1}(t_0 \sec \beta' / s_0). \qquad \ldots\ldots(36)$$

It will be noted that $\sec \beta'$ appears in (35), (36) whereas the equivalent frequency s and direction ψ appropriate to experimental work are the values obtained previously, namely

$$s = (s_0^2 + t_0^2)^{1/2}, \qquad \psi = \tan^{-1}(t_0/s_0). \qquad \ldots\ldots(37)$$

This is because in experimental work the object and image will lie in planes perpendicular to the optical axis, whereas the analytical expression (34) has the form, apart from the correction term, which is appropriate to the case of object and image planes perpendicular to the principal ray. Indeed, it follows from the comparison of the first term of (34) and (31), that the response is the same for the frequencies $\{s = \sqrt{(s_0^2 + t_0^2 \sec^2 \beta')},\ \psi = \tan^{-1}(t_0/s_0 \cos \beta')\}$ and $\{s = \sqrt{(s_0^2 + t_0^2)},\ \psi = \tan^{-1}(t_0/s_0)\}$ in the two cases of planes perpendicular to the principal ray and the optical axis respectively, provided that the correction term in (34) is neglected. It appears from this that the equivalent structure in the former case is that of which $\{s = \sqrt{(s_0^2 + t_0^2)},\ \psi = \tan^{-1}(t_0/s_0)\}$ is the projection, in both spacing and direction, on the image plane perpendicular to the optical axis.

For an aberration-free system, having a pupil symmetrical about the axes of x and y, the pupil function $f(x, y)$ is equal to unity within A and is zero outside. The second integral in (34) then vanishes, since the integrand is an odd function in both x and y. Thus there is no correction to add when the optical system has no aberration. In other cases, we may note that the constant q will not exceed 1/16, even in an extreme case. Furthermore, the second integral in (34) being a 'moment' of the integrand in the first integral, with $|x| < 1$ and $|y| < 1$, the modulus of the second integral is unlikely to exceed that of the first. The second integral will therefore always be a small correction, which in practical cases will be less than a few per cent. In considering the general properties of the response function $D(s, \psi)$, we shall accordingly neglect this correction.

In contrast to this, the presence of $\sec \beta'$ in the equivalent frequency s in (35) may lead to a significant difference between the response calculated for a given frequency in the image plane perpendicular to the principal ray (which is usually implied in calculations of this kind) and that for the same frequency in the image plane perpendicular to the optical axis, which is usually the case of practical importance. When $\beta' = 30°$, for example, the single frequency equivalent to $(0, t_0)$ is $s = t_0 \sec \beta' = 1{\cdot}15 t_0$. When allowance is also made for the smaller value of the angular aperture α', compared with that obtaining for the axial image point, an additional factor $\sec \beta'$ is involved in determining t_0 from the wave number R', if the value of α' used is that corresponding to the axial F-ratio of the optical system. Thus, with the usual treatment employing the aperture angle of the axial pencil and the approximation $\cos \beta' = 1$, to find the response for $R' = 20$ lines/mm at a field angle of 30° in the actual image plane a value of s would need to be employed (in calculations using (31)) corresponding to approximately 26 lines/mm. For most practical systems such a difference would be very significant.

In what follows, we shall ignore the correction term in (34), and merely bear in mind that, in analytical work, the equivalent frequency s is that given by the expression (35), and the direction of the equivalent unidimensional structure is given by (36). The response is then given by (31), using the appropriately rotated pupil function.

§ 3. General Properties of the Response Function $D(s)$

If the optical system is free from aberration and perfectly focused, $W(x,y)=0$ and $f(x,y)=\tau(x,y)$. In this case

$$D_0(s,\psi)=\frac{1}{A}\iint_{-\infty}^{+\infty}\tau(x+\tfrac{1}{2}s,y)\tau(x-\tfrac{1}{2}s,y)\,dx\,dy \qquad \ldots\ldots(38)$$

and $D_0(s,\psi)$ is therefore necessarily real and positive. There will, in consequence, be no phase-shift or reversal of contrast. Usually the pupil may be taken to be of uniform transparency with sufficient accuracy. $D(s,\psi)$ is then simply equal to the ratio of the area common to two pupils, centred respectively on the points $(\pm\tfrac{1}{2}s,0)$, to that of the pupil itself. It follows that in most cases $D(s,\psi)$ decreases monotonically from unity as s increases from zero.† This is in contrast to what happens in coherent light, in which case the response is equal to unity for the whole range of resolved frequencies. For circular and rectangular pupils the response functions have the well-known forms

$$\left.\begin{aligned} D_0(s)&=\frac{1}{\pi}(2\beta-\sin 2\beta) \qquad \beta=\cos^{-1}|\tfrac{1}{2}s| \\ D_0(s)&=1-|\tfrac{1}{2}s| \end{aligned}\right\} \qquad \ldots\ldots(39)$$

for $|s|\leqslant 2$, and they are both zero for $|s|>2$. In these expressions ψ has been omitted: in the former, $D_0(s)$ is independent of ψ, and, in the latter, the structure is assumed to lie parallel with sides of the rectangular aperture.

If the optical system has aberration, in which must be included any defect of focus, the response is given by

$$D(s,\psi)=\frac{1}{A}\iint_{-\infty}^{+\infty}\tau(x+\tfrac{1}{2}s,y)\tau(x-\tfrac{1}{2}s,y)\exp\{ik[W(x+\tfrac{1}{2}s,y)-W(x-\tfrac{1}{2}s,y)]\}\,dx\,dy. \qquad \ldots\ldots(40)$$

Since $\tau(x,y)$ is real and positive, the integrand of (38) is the modulus of that in (40). It follows that

$$|D(s,\psi)|\leqslant D_0(s) \qquad \ldots\ldots(41)$$

and the presence of aberration will therefore always result in a decrease in contrast for every spatial frequency other than zero. There may, in addition, be a phase-shift or reversal of contrast for any given frequency.

We shall now denote the region in the (x,y) plane common to two pupils centred respectively on the points $(\pm\tfrac{1}{2}s,0)$ by the symbol S. In place of (40) we now write

$$D(s,\psi)=\frac{1}{A}\iint_S\tau(x+\tfrac{1}{2}s,y)\tau(x-\tfrac{1}{2}s,y)\exp\{iksW(x,y\,;\,s)\}. \qquad \ldots\ldots(42)$$

in which

$$sW(x,y\,;\,s)=W(x+\tfrac{1}{2}s,y)-W(x-\tfrac{1}{2}s,y). \qquad \ldots\ldots(43)$$

A useful expression for $W(x,y\,;\,s)$ is obtained by expanding each of the terms on the right-hand side of (43) in a Taylor series about the point (x,y), giving

$$W(x,y\,;\,s)=W_x'(x,y)+\frac{1}{3!}\left(\frac{s}{2}\right)^2W_x'''(x,y)+\ldots \qquad \ldots\ldots(44)$$

---

† An aperture consisting of two separate slits is an example where this is not the case (Baker 1955).

In this series the number of terms will be finite, and will rarely exceed three, since $W(x,y)$ is usually at most of the sixth degree in x, if we exclude microscope objectives, which are in any case seldom used in incoherent light.

The value of $D(s,\psi)$ will depend on the variation of $W(x,y;\,s)$ over the region of integration. It is at once apparent from (44) that for larger aberrations $D(s,\psi)$ will only be of the same order as $D_0(s)$, the response in the absence of aberration, when s is sufficiently small. In other words, the larger the aberration the smaller the range of resolved frequencies. However, it is the ξ' component of the transverse ray aberration, which is proportional to $W_x'(x,y)$, and its successive even derivatives which determine the quality of the image, and not the wave aberration, $W(x,y)$, directly. This is again in marked contrast to the case of image formation in coherent light, for which the response for the frequency pair (x,y) is $\exp ikW(x,y)$. An immediate consequence of this result is that astigmatism has no effect on the image of a line structure set perpendicular to one of the astigmatic axes, and viewed in a plane passing through the corresponding astigmatic line focus.

We shall now show that the s-derivative of $|D(s,\psi)|$ at the origin $s=0$ is unaffected by the presence of aberration. From (42) and (44),

$$D(s,\psi)=\frac{1}{A}\iint_S \tau(x+\tfrac{1}{2}s,y)\tau(x-\tfrac{1}{2}s,y)\{1+iksW_x'(x,y)+\mathrm{O}(s^2)\}dxdy$$

or, if $D_0(s)$ is again used to denote the response in the absence of aberration,

$$D(s,\psi)=D_0(s)+\frac{iks}{A}\iint_S W_x'(x,y)\tau(x+\tfrac{1}{2}s,y)\tau(x-\tfrac{1}{2}s,y)dxdy+\mathrm{O}(s^2). \qquad \ldots.(45)$$

It follows at once that

$$\left\{\frac{\partial}{\partial s}\,|D(s,\psi)|\right\}_{s=0}=\left\{\frac{\partial}{\partial s}\,D_0(s)\right\}_{s=0} \qquad \ldots\ldots(46)$$

but, also, that the argument of $D(s,\psi)$, corresponding to a phase shift, may involve a term in s. The result (46) would seem to explain the practical fact that recognizable images of coarse structures are obtained even with very large aberrations present in the image-forming system. For a system with a circular pupil, and $\tau(x,y)=1$, it follows from (39) that $D_s'(0,\psi)=-2/\pi$.

In many cases the s-derivative of $D(s,\psi)$ is also independent of aberration at the limit of resolution. In particular this is true for a circular pupil, when the derivative is zero.

To avoid unnecessary discussion, we shall now suppose the image-forming wave fronts to be of uniform amplitude, and the pupil to be circular. This requires $\tau(x,y)=1$, and the region of integration in (42) is that common to two circles of unit radius centred respectively on the two points $(\pm\tfrac{1}{2}s,0)$ in the (x,y) plane. This region of integration we shall denote as before by S. It is important in much of what follows that this region is symmetrical about the axes of both x and y.

If the aberration $W(x,y)$, after any rotation of axes to take account of the inclination ψ of the line structure under consideration to the original axis y_0, is an even function of x, the function $W(x,y;\,s)$ defined by (43) will be an odd function of x. Thus, a typical example will be spherical aberration, for which

$$W(x,y)=W_{20}(x^2+y^2)+W_{40}(x^2+y^2)^2+W_{60}(x^2+y^2)^3$$

the various terms here representing a defect of focus, primary and secondary spherical aberration respectively. Each of the odd derivatives of $W(x,y)$

contains x as a factor, the remaining factors being even polynomials in x. For such cases, the region of integration also being symmetrical about the y axis, (42) reduces to

$$D(s,\psi)=\frac{1}{\pi}\iint_S \cos\{ksW(x,y\,;\,s)\}dxdy \qquad \ldots\ldots(47)$$

the area of the pupil A being now put equal to π. Hence, since $D(s,\psi)$ is wholly real, there is no phase shift with even types of aberration, other than $\pm\pi$, which correspond to reversal of contrast.

If the aberration function is an odd function of x, its odd derivatives will give rise to certain terms in $sW(x,y\,;\,s)$ which are independent of (x,y), and each containing a power of s as factor. The exponential function of this polynominal in s may be taken outside the integral sign in (42). It is of unit modulus, and therefore represents a pure phase shift which, in general, is a function of the frequency s. Moreover, the remainder of $W(x,y\,;\,s)$ will have both odd and even terms in x, so that the integral itself will usually be complex. Hence, with odd types of aberration, such as coma, there will usually be both phase shift and reduction of contrast.

The form of the integrand in (42) leads one to expect $|D(s,\psi)|$ to resemble the modulus of a damped sine curve. Analytical forms of $D(s,\psi)$ showing this behaviour have recently been published for an aberration-free lens with defect of focus (Hopkins 1955) and for an optical system suffering from astigmatism (De 1956). Some of the former results have been verified experimentally (Baker 1955).

In many investigations it is more convenient to use the relative modulation $M(s)\exp\{i\theta(s)\}$, defined by

$$M(s)\exp\{i\theta(s)\}=D(s,\psi)/D_0(s). \qquad \ldots\ldots(48)$$

The modulus $M(s)$ is the ratio of the contrast reduction in the presence of aberration to that obtaining with the same system in the absence of aberration: the argument $\theta(s)$, being identical with that of $D(s,\psi)$, denotes directly the phase-shift. The relative modulation factor $M(s)\exp\{i\theta(s)\}$ is given by the expression

$$M(s)\exp\{i\theta(s)\}=\iint_S \exp\{iksW(x,y\,;\,s)\}dxdy \Big/ \iint_S dxdy \qquad \ldots\ldots(49)$$

for the case of image-forming wave fronts of uniform amplitude.

In many cases of practical interest only small values of s ($<0{\cdot}2$, say) come in question. If, for these cases, we neglect the terms in (44) involving the square and higher powers of s (49) reduces to

$$M(s)\exp\{i\theta(s)\}=\iint_S \exp\{iksW_x'(x,y)\}dxdy \Big/ \iint_S dxdy.$$

When s is small the difference between the ratio of the values of these two integrations carried out over the whole unit circle centred on the origin and over the region S will be small. It follows that $M(s)$ will be practically unaltered if, for an increased value of s, the aberration coefficients in $W_x'(x,y)$ are reduced by the same factor. Hence, if the relative response $M(s)\exp\{i\theta(s)\}$ is known for any given optical system, the response for a system of identical construction but of k times the focal length will be $M(ks)\exp\{i\theta(ks)\}$. If, for example, the focal length is doubled the same response will be obtained for frequencies having half the original values. The importance in practical optics of this rule will be evident.

We shall conclude by showing, with reference to a particular case, that it is in general not possible to find the frequency response of a combination of optical systems from a knowledge of their separate response functions. If there are two optical systems of which the aberration functions are $W_2(x,y)$, $W_1(x,y)$, both being even functions of x and y, and $W_2(x,y) = -W_1(x,y)$, the responses $D_1(s)$, $D_2(s)$ are identical. For, in this case, $W_2(x,y;s) = -W_1(x,y;s)$, and by (47) these will lead to identical values for the response. It is thus not possible from the response curves to distinguish between two systems having equal but opposite symmetrical aberration. If now two systems having aberrations $\pm W(x,y)$ are combined, the resulting system will be free from aberration and have a response equal to $D_0(s)$. If the two systems have identical aberrations of the same sign, however, the combined system will have an aberration equal to $2W(x,y)$, apart from a small cross-term, and the response will be necessarily less than $D_0(s)$. It follows at once that the response of the combined system cannot, in general, be found from the responses of the separate components.

The above result has bearing on the measurement of the response of an optical system in practice. If, for example, the system is arranged to form an image of a unidimensional structure, which image is subsequently enlarged for convenience of photometric scanning, the enlarging lens must be free from aberration, since no correction can be made for the effects of any departure of the response of this lens system from that obtained in the absence of aberration.

REFERENCES

BAKER, L. R., 1955, *Proc. Phys. Soc.* B, **68,** 871.
DE, M., 1956, *Proc. Roy. Soc.* A, **233,** 91.
DUFFIEUX, P. M., 1946, *L'integrale de Fourier et ses applications à l'optique* (Besançon: private publication).
HOPKINS, H. H., 1949, *Proc. Phys. Soc.* B, **62,** 22; 1950, *Wave Theory of Aberrations* (Oxford: Clarendon Press); 1953, *Proc. Roy. Soc.* A, **217,** 408; 1955, *Ibid.*, **231,** 91.
HOPKINS, R. E., 1954, Symposium on Optical Image Evaluation 1951, Circ. U.S. Bur. Stand. 526.
RAYLEIGH, LORD, 1896, *Phil. Mag.*, **42,** 167.

Radiative Transfer with Central Sources of Non-Uniform Directional Intensity

By J. T. JEFFERIES

Division of Physics, National Standards Laboratory, Commonwealth Scientific and Industrial Research Organization, Sydney

Communicated by G. H. Briggs ; MS. received 21st December 1955

Abstract. Approximate expressions are obtained for the total intensity in a spherically or cylindrically symmetric diffuser containing respectively a point or a line source centrally placed, the source emitting with an arbitrary directional dependence. Results are given for homogeneous media and for simple cases of composite media.

---

§ 1. Introduction

In an earlier publication (Jefferies 1955) an approximation was obtained for the total intensity of radiation in a uniform infinite isotropic diffusing medium with a point or a line source emitting uniformly in all directions. The method there developed is now extended to the case of a non-isotropic source and the analysis then applied to some simple cases of symmetric composite models.

The total intensity J at a point in the medium is defined in terms of the specific intensity I at that point by the equation

$$J=\int I d\omega \qquad \ldots\ldots(1)$$

the integration extending over a solid angle 4π. Using an approximation due to Eddington, the net flux $\mathbf{F}$ per unit area in the medium may be written in the vector form

$$\nabla J=-3\kappa\mathbf{F} \qquad \ldots\ldots(2)$$

where κ is the extinction coefficient and the total intensity of the diffuse component of the radiation in a uniform scattering medium is then given by the equation

$$\nabla^2 J=3\kappa^2(\lambda J-4\pi P) \qquad \ldots\ldots(3)$$

where $\lambda=1-\varpi_0$, ϖ_0 being the albedo for single scattering and P a source function defined as the ratio of the rate of emission of diffuse radiation per unit volume and solid angle to the extinction coefficient κ. The source function takes account of the scattering of direct radiation from sources inside or outside the medium.

In §§ 2 and 3 we consider the total intensity of radiation in an infinite homogeneous diffusing medium illuminated by a point or a line source which emits with an arbitrary angular dependence; §§ 4 and 5 deal with composite models having different properties inside and outside a sphere and cylinder, respectively, concentric with isotropic sources, an extension to non-isotropic sources being given in § 6. The extension to a number of such concentric media is straightforward and is not considered in detail.

§ 2. POINT SOURCE ILLUMINATION

Consider a point source in an infinite scattering medium and suppose the source emits energy at a rate $4\pi Q$ with a polar distribution $f(\theta, \phi)$, θ and ϕ being angular coordinates of a spherical polar system. Expanding $f(\theta, \phi)$ in a series of surface harmonics $V_n(\theta, \phi)$† in the form

$$f(\theta, \phi) = \sum_{n=0}^{\infty} V_n(\theta, \phi) \qquad \ldots\ldots(4)$$

the source function may be written

$$P(r, \theta, \phi) = \frac{Q\varpi_0\, e^{-\kappa r}}{4\pi r^2} \sum V_n(\theta, \phi) \qquad \ldots\ldots(5)$$

so that the transfer equation (3) becomes, writing $\mu = \cos\theta$,

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial J}{\partial r}\right) + \frac{1}{r^2}\frac{\partial}{\partial \mu}\left\{(1-\mu^2)\frac{\partial J}{\partial \mu}\right\} + \frac{1}{r^2(1-\mu^2)}\frac{\partial^2 J}{\partial \phi^2} = 3\kappa^2\left[\lambda J - \frac{Q\varpi_0\, e^{-\kappa r}}{r^2}\sum V_n(\theta, \phi)\right]. \qquad \ldots\ldots(6)$$

Assuming a solution for J of the form

$$J = \sum_{n=0}^{\infty} \mathcal{J}_n(r) V_n(\theta, \phi) \qquad \ldots\ldots(7)$$

and substituting in (6) we obtain, after some simplification,

$$\sum V_n(\theta, \phi)\left\{\frac{d^2\mathcal{J}_n}{dt^2} + \frac{2}{t}\frac{d\mathcal{J}_n}{dt} - \left[1 + \frac{n(n+1)}{t^2}\right]\mathcal{J}_n\right\} = -C\frac{e^{-\beta t}}{t^2}\sum V_n(\theta, \phi) \quad \ldots(8)$$

where $t = \sqrt{(3\lambda)}\kappa r$, $\beta = 1/\sqrt{(3\lambda)}$ and $C = 3Q\kappa^2\varpi_0$.

The functions $\mathcal{J}_n$ are then given by

$$\frac{d^2\mathcal{J}_n}{dt^2} + \frac{2}{t}\frac{d\mathcal{J}_n}{dt} - \left[1 + \frac{n(n+1)}{t^2}\right]\mathcal{J}_n = -\frac{C\, e^{-\beta t}}{t^2} \qquad \ldots\ldots(9)$$

which has a solution

$$\mathcal{J}_n(t) = C\left[t^{-1/2}\mathrm{K}_{n+1/2}(t)\int_{a_n}^{t} x^{-1/2}\mathrm{I}_{n+1/2}(x)\,dx - t^{-1/2}\mathrm{I}_{n+1/2}(t)\int_{b_n}^{t} x^{-1/2}\mathrm{K}_{n+1/2}(x)\,dx\right] \qquad \ldots\ldots(10)$$

where the integration constants a_n and b_n are determined by the boundary conditions.

A first condition is that, as $t \to \infty$, $J \to 0$ for all values of θ and ϕ and, since $\mathrm{I}_{n+1/2}(t) \to \infty$ with increasing t, this requires that $b_n = \infty$ for all n.

A second boundary condition may be obtained as follows. The emission into a solid angle $d\omega$ from a volume dV of the medium at the point with spherical polar coordinates (ξ, η, χ) is

$$\kappa\left[\frac{\varpi_0 J(\xi, \eta, \chi)}{4\pi} + P(\xi, \eta, \chi)\right] d\omega\, dV.$$

The net radial flux through unit area at (r, θ, ϕ) is then

$$F(r, \theta, \phi) = \int \kappa\left[\frac{\varpi_0 J(\xi, \eta, \chi)}{4\pi} + P(\xi, \eta, \chi)\right] \exp\left(-\kappa\,|r - \xi|\right)\Delta\omega\, dV \quad \ldots\ldots(11)$$

† This may be done for a large class of functions $f(\theta, \phi)$, see for example, Hobson (1931).

where $\Delta\omega$ is the solid angle subtended at (ξ, η, χ) by the unit area at (r, θ, ϕ). Equation (11) gives a first integral of the transfer equation since, from (2), the radial component F_r of the net flux is $-(1/3\kappa)(\partial J/\partial r)$. The integral in (11) may be evaulated in the limit as $r \rightarrow 0$ and comparison with the corresponding limit of $-(1/3\kappa)(\partial J/\partial r)$ obtained from (10) shows that $a_n = 0$ for all n. This is equivalent to the condition that, as $r \rightarrow 0$, the net radial flux is independent of θ and ϕ.

Thus if a point source of strength $4\pi Q$ and polar distribution $f(\theta, \phi) = \Sigma V_n(\theta, \phi)$ is situated in an infinite scattering medium, the total intensity J at a point (r, θ, ϕ) is given by the equation

$$J = \sum \mathcal{J}_n(t) V_n(\theta, \phi) \qquad \ldots\ldots(12)$$

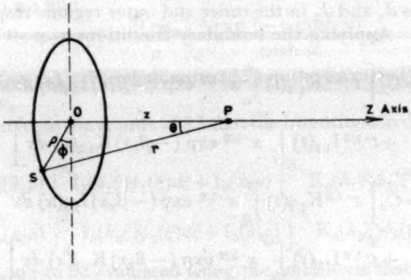
where $\mathcal{J}_n(t) =$

$$C\left[t^{-1/2} \mathrm{K}_{n+1/2}(t) \int_0^t x^{-1/2} \mathrm{e}^{-\beta x} \mathrm{I}_{n+1/2}(x)\, dx + t^{-1/2} \mathrm{I}_{n+1/2}(t) \int_t^\infty x^{-1/2} \mathrm{e}^{-\beta x} \mathrm{K}_{n+1/2}(x)\, dx\right].$$

§ 3. LINE SOURCE ILLUMINATION

Consider an infinite line source emitting energy at a rate $4\pi Q$ per unit length and suppose the source to be the z axis of a system of cylindrical coordinates. If the source emits with an angular distribution $f(\theta, \phi)$ where the angles θ and ϕ are as shown in the figure, the source function at the point S is

$$P(\rho, \phi) = \int_{-\infty}^{+\infty} \frac{Q \varpi_0\, \mathrm{e}^{-\kappa r} f(\theta, \phi)\, dz}{4\pi r^2}. \qquad \ldots\ldots(13)$$



If $P(\rho, \phi)$ is expanded as a Fourier series of the form

$$4\pi P(\rho, \phi) = \frac{F(\rho)}{\rho} \sum (a_n \cos n\phi + b_n \sin n\phi) \qquad \ldots\ldots(14)$$

and assuming J may be expanded as

$$J = \sum \mathcal{J}_n(\rho)(a_n \cos n\phi + b_n \sin n\phi) \qquad \ldots\ldots(15)$$

the transfer equation (3) shows that the function $\mathcal{J}_n(\rho)$ is given by

$$\frac{d^2\mathcal{J}_n}{dv^2} + \frac{1}{v}\frac{d\mathcal{J}_n}{dv} - \left(1 + \frac{n^2}{v^2}\right)\mathcal{J}_n = \frac{\psi(v)}{v} \qquad \ldots\ldots(16)$$

where $v=\sqrt{(3\lambda)}\kappa\rho$ and $\psi(v)=\kappa\sqrt{(3/\lambda)}F(v/\sqrt{(3\lambda)}\kappa)$. Equation (16) can be solved by variation of parameters to give

$$\mathcal{J}_n(v)=\mathrm{K}_n(v)\int_{a_n}^{v}\mathrm{I}_n(x)\psi(x)\,dx-\mathrm{I}_n(v)\int_{b_n}^{v}\mathrm{K}_n(x)\psi(x)\,dx \quad \ldots\ldots(17)$$

where a_n and b_n are to be assigned in accordance with the boundary conditions. Using methods analogous to those given above for the case of a point source, it follows that $a_n=0$ and $b_n=\infty$, so that the solution for the total intensity J is given by (15) with the functions $\mathcal{J}_n(v)$ given by

$$\mathcal{J}_n(v)=\mathrm{K}_n(v)\int_{0}^{v}\mathrm{I}_n(x)\psi(x)\,dx+\mathrm{I}_n(v)\int_{v}^{\infty}\mathrm{K}_n(x)\psi(x)\,dx. \quad \ldots\ldots(18)$$

§ 4. THE DISCONTINUOUS SPHERICALLY SYMMETRIC MODEL

Consider a point source emitting radiation uniformly in an isotropically scattering medium which is separated, by a spherical surface of radius r_0, into two homogeneous media with different properties. We denote quantities inside and outside the sphere with subscript 1 and 2 respectively.

The intensity in the medium is given by the solution of the equation

$$\nabla^2 J=3\kappa^2(\lambda J-4\pi P) \quad \ldots\ldots(19)$$

with

$$P_1=\frac{Q\varpi_1\exp(-\kappa_1 r)}{4\pi r^2} \quad \text{and} \quad P_2=\frac{Q\varpi_2\exp\{-(\kappa_1-\kappa_2)r_0\}\exp(-\kappa_2 r)}{4\pi r^2}.$$

The intensities J_1 and J_2 in the inner and outer regions respectively are given by (10) for $n=0$. Applying the boundary conditions as $r\to 0$ and $r\to\infty$ we find

$$\begin{aligned}J_1(t)=C_1\Big[&t^{-1/2}\mathrm{K}_{1/2}(t)\int_0^t x^{-1/2}\exp(-\beta_1 x)\mathrm{I}_{1/2}(x)\,dx\\&+t^{-1/2}\mathrm{I}_{1/2}(t)\int_t^T x^{-1/2}\exp(-\beta_1 x)\mathrm{K}_{1/2}(x)\,dx\Big]\end{aligned} \quad \ldots\ldots(20)$$

$$\begin{aligned}J_2(s)=C_2\Big[&s^{-1/2}\mathrm{K}_{1/2}(s)\int_S^s x^{-1/2}\exp(-\beta_2 x)\mathrm{I}_{1/2}(x)\,dx\\&+s^{-1/2}\mathrm{I}_{1/2}(s)\int_s^\infty x^{-1/2}\exp(-\beta_2 x)\mathrm{K}_{1/2}(x)\,dx\Big]\end{aligned} \quad \ldots\ldots(20')$$

where $C_1=3\kappa_1{}^2Q\varpi_1$, $C_2=3\kappa_2{}^2Q\varpi_2\exp[-(\kappa_1-\kappa_2)r_0]$, $t=\sqrt{(3\lambda_1)}\kappa_1 r$, $s=\sqrt{(3\lambda_2)}\kappa_2 r$ and the integration constants T and S are to be fitted by the continuity conditions across the spherical surface at $r=r_0$. The requirement that the total intensity and the net flux are continuous yields the conditions that at $r=r_0$

$$J_1(t)=J_2(s) \quad \ldots\ldots(21), \qquad \beta_1^{-1}\frac{\partial J_1}{\partial t}=\beta_2^{-1}\frac{\partial J_2}{\partial s} \quad \ldots\ldots(21')$$

and these conditions may be used immediately with (20) and (20′) to obtain values for the integrals

$$Y_1=\int_{t_0}^{T}x^{-1/2}\exp(-\beta_1 x)\mathrm{K}_{1/2}(x)\,dx, \qquad X_2=\int_{S}^{s_0}x^{-1/2}\exp(-\beta_2 x)\mathrm{I}_{1/2}(x)\,dx.$$

The solutions J_1 and J_2 can then be written in terms of X_2 and Y_1 in the form

$$J_1(t)=C_1\left\{t^{-1/2}\mathrm{K}_{1/2}(t)\int_0^t x^{-1/2}\exp(-\beta_1x_1)\,\mathrm{I}_{1/2}(x)\,dx\right.$$
$$\left.+t^{-1/2}\mathrm{I}_{1/2}(t)\left[\int_t^{t_0} x^{-1/2}\exp(-\beta_1x)\,\mathrm{K}_{1/2}(x)\,dx+Y_1\right]\right\} \quad \ldots\ldots(22)$$

$$J_2(s)=C_2\left\{s^{-1/2}\mathrm{K}_{1/2}(s)\left[\int_{s_0}^{s} x^{-1/2}\exp(-\beta_2x)\,\mathrm{I}_{1/2}(x)\,dx+X_2\right]\right.$$
$$\left.+s^{-1/2}\mathrm{I}_{1/2}(s)\int_s^{\infty} x^{-1/2}\exp(-\beta_2x)\,\mathrm{K}_{1/2}(x)\,dx\right\}. \quad \ldots\ldots(22')$$

The extension of the above results to three or more concentric media is straightforward.

Examples of some problems which may be treated using the above results are:

(i) $\kappa_2=0$ (finite sphere in a vacuum),
(ii) $\kappa_2=\infty$ (reflecting surface on a finite sphere),
(iii) $\kappa_1=0$ (hole in an infinite diffuser).

§ 5. THE DISCONTINUOUS CYLINDRICALLY SYMMETRIC MODEL

Consider a line source emitting radiation uniformly in an isotropically scattering medium which has constant properties inside and outside a cylinder of radius r_0. We distinguish quantities inside and outside the cylinder respectively with suffices 1 and 2.

In the notation of § 3 the source function in region (1) is readily shown to be

$$4\pi P_1(\rho)=2\underset{\sim}{Q}\varpi_1\int_0^{\infty}\frac{\exp(-\kappa_1r)}{r^2}\,dz=\frac{2\underset{\sim}{Q}\varpi_1}{\rho}\int_0^{\infty}\frac{\exp(-\kappa_1\rho\cosh u)}{\cosh u}\,du \quad \ldots(23)$$

which may be written $C_1\psi_1(\rho)/3\kappa_1^{\,2}\rho$, where $C_1=6\kappa_1^{\,2}\underset{\sim}{Q}\varpi_1$.

For region (2) we find

$$4\pi P_2(\rho)=\frac{2\underset{\sim}{Q}\varpi_2}{\rho}\int_0^{\infty}\frac{\exp(-\gamma\cosh u)}{\cosh u}\,du \quad \ldots\ldots(23')$$

with $\gamma=\kappa_1\rho_0+\kappa_2(\rho-\rho_0)$. We shall write (23′) in the form $4\pi P_2(\rho)=C_2\psi_2(\rho)/3\kappa_2^{\,2}\rho$ with $C_2=6\kappa_2^{\,2}\underset{\sim}{Q}\varpi_2$.

It follows then from the results of § 3 that the intensities in the two regions may be written

$$J_1(\rho)=C_1\left[\mathrm{K}_0(k_1\rho)\int_0^{\rho}\mathrm{I}_0(k_1\xi)\psi_1(\xi)d\xi+\mathrm{I}_0(k_1\rho)\int_{\rho}^{A_1}\mathrm{K}_0(k_1\xi)\psi_1(\xi)\,d\xi\right] \quad \ldots\ldots(24)$$

$$J_2(\rho)=C_2\left[\mathrm{K}_0(k_2\rho)\int_{A_2}^{\rho}\mathrm{I}_0(k_2\xi)\psi_2(\xi)d\xi+\mathrm{I}_0(k_2\rho)\int_{\rho}^{\infty}\mathrm{K}_0(k_2\xi)\psi_2(\xi)\,d\xi\right] \quad \ldots(24')$$

where A_1 and A_2 are to be evaluated using the conditions that, across the surface $\rho=\rho_0$, total intensity and the radial flux are continuous. These conditions give values for the integrals

$$W_1=\int_{\rho_0}^{A_1}\mathrm{K}_0(k_1\xi)\psi_1(\xi)\,d\xi, \qquad Z_2=\int_{A_2}^{\rho_0}\mathrm{I}_0(k_2\xi)\psi_2(\xi)\,d\xi$$

in terms of which the solutions (24, 24′) may be written:

$$J_1(\rho)=C_1\left\{\mathrm{K}_0(k_1\rho)\int_0^{\rho}\mathrm{I}_0(k_1\xi)\psi_1(\xi)\,d\xi+\mathrm{I}_0(k_1\rho)\left[\int_{\rho}^{\rho_0}\mathrm{K}_0(k_1\xi)\psi_1(\xi)\,d\xi+W_1\right]\right\}. (25)$$

$$J_2(\rho)=C_2\left\{\mathrm{K}_0(k_2\rho)\left[\int_{\rho_0}^{\rho}\mathrm{I}_0(k_2\xi)\psi_2(\xi)\,d\xi+Z_2\right]+\mathrm{I}_0(k_2\rho)\int_{\rho}^{\infty}\mathrm{K}_0(k_2\xi)\psi_2(\xi)d\xi\right\}. (25')$$

The extension to models with two or more coaxial discontinuous cylindrical surfaces is again straightforward.

§ 6. NON-ISOTROPIC SOURCES

If a point source of radiation in a discontinuous model of § 4 radiates non-isotropically with a polar distribution $f(\theta, \phi)=\Sigma V_n(\theta, \phi)$, the total intensity may be found by methods similar to those of § 2 for the homogeneous model. In the notation used there, we may write

$$J_1(t)=\sum \mathcal{J}_n^{(1)}(t)V_n(\theta, \phi) \quad \ldots\ldots(26) \qquad J_2(t)=\sum \mathcal{J}_n^{(2)}(s)V_n(\theta, \phi). \quad \ldots\ldots(26')$$

The $\mathcal{J}_n$ in (26, 26′) are given by equation (10) with $C_1=3\kappa_1^2Q\varpi_1$, and $C_2=3\kappa_2^2Q\varpi_2, \exp[-(\kappa_1-\kappa_2)r_0]$. The conditions near the source and at infinity give $a_n^{(1)}=0$ and $b_n^{(2)}=\infty$ as for the homogeneous model. Across the spherical boundary the total intensity and the radial component of the net flux must be continuous for all values of θ, which requires that, at $r=r_0$,

$$\mathcal{J}_n^{(1)}(t)=\mathcal{J}_n^{(2)}(s) \quad \ldots\ldots(27), \qquad \beta_1^{-1}\frac{d\mathcal{J}_n^{(1)}(t)}{dt}=\beta_2^{-1}\frac{d\mathcal{J}_n^{(2)}(s)}{ds} \quad \ldots\ldots(27')$$

for all values of n.

Application of these conditions as in § 4 gives values for the integrals

$$Y_1^{(n)}=\int_{t_0}^{T_n} x^{-1/2}\exp(-\beta_1 x)\mathrm{K}_{n+1/2}(x)\,dx, \qquad X_2^{(n)}=\int_{S_n}^{s_0} x^{-1/2}\exp(-\beta_2 x)\mathrm{I}_{n+1/2}(x)\,dx$$

so that the $\mathcal{J}_n$ become

$$\begin{aligned}\mathcal{J}_n^{(1)}=C_1\Big\{&t^{-1/2}\mathrm{K}_{n+1/2}(t)\int_0^t x^{-1/2}\exp(-\beta_1 x)\mathrm{I}_{n+1/2}(x)\,dx\\ &+t^{-1/2}\mathrm{I}_{n+1/2}(t)\left[\int_t^{t_0} x^{-1/2}\exp(-\beta_1 x)\mathrm{K}_{n+1/2}(x)\,dx+Y_1^{(n)}\right]\Big\} \quad \ldots\ldots(28)\end{aligned}$$

$$\begin{aligned}\mathcal{J}_n^{(2)}=C_2\Big\{&s^{-1/2}\mathrm{K}_{n+1/2}(s)\left[\int_{s_0}^{s} x^{-1/2}\exp(-\beta_2 x)\mathrm{I}_{n+1/2}(x)\,dx+X_2^{(n)}\right]\\ &+s^{-1/2}\mathrm{I}_{n+1/2}(s)\int_s^\infty x^{-1/2}\exp(-\beta_2 x)\mathrm{K}_{n+1/2}(x)\,dx\Big\}. \quad \ldots\ldots(28')\end{aligned}$$

For a non-isotropic line source situated in the discontinuous model of § 5, the treatment is completely analogous to the above. We find, in the notation of § 3:

$$J_i=\sum \mathcal{J}_n^{(i)}(\rho)[a_n\cos n\phi+b_n\sin n\phi]$$

where

$$\mathcal{J}_n^{(1)}(\rho)=C_1\left\{\mathrm{K}_n(k_1\rho)\int_0^\rho \mathrm{I}_n(k_1\xi)\psi_1(\xi)\,d\xi+\mathrm{I}_n(k_1\rho)\left[\int_\rho^{\rho_0}\mathrm{K}_n(k_1\xi)\psi_1(\xi)\,d\xi+W_1^{(n)}\right]\right\} \quad \ldots\ldots(29)$$

$$\mathcal{J}_n^{(2)}(\rho)=C_2\left\{\mathrm{K}_n(k_2\rho)\left[\int_{\rho_0}^{\rho}\mathrm{I}_n(k_2\xi)\psi_2(\xi)\,d\xi+Z_2^{(n)}\right]+\mathrm{I}_n(k_2\rho)\int_\rho^\infty \mathrm{K}_n(k_2\xi)\psi_2(\xi)\,d\xi\right\} \quad \ldots\ldots(29')$$

$W_1^{(n)}$ and $Z_2^{(n)}$ being integrals whose values are obtained from the continuity of total intensity and net flux across the cylindrical surface of discontinuity.

REFERENCES

HOBSON, E. W., 1931, *The Theory of Spherical and Ellipsoidal Harmonics* (Cambridge : University Press).

JEFFERIES, J. T., 1955, *Optica Acta*, **2**, 109.

RESEARCH NOTES

Disorientation of Single Crystal Substrate due to Epitaxial Stress

By A. GOSWAMI

National Chemical Laboratory of India, Poona

MS. received 28th December 1955

When a crystal grows epitaxially over another, a lattice stress is usually developed at the substrate–deposit interface. The origin of this stress is not difficult to see. The deposit atoms on their arrival at the substrate surface tend to take up positions at potential hollows and unless the lattice positions of the deposit atoms coincide appropriately with the substrate potential field distributions the deposit lattice will be in a strained condition. If this strain exceeds a certain limit depending upon the nature and structure of the deposit, it will be relieved by disorientation or twinning of the crystals or by rotational slip. Many examples are now available which clearly illustrate the above process of relief of interfacial stress in the epitaxial overgrowths.

The formation of twinned crystals in electrodeposits (Cochrane 1936, Wilman 1940, Finch, Wilman and Yang 1947) and thin metal films deposited in vacuum (Lassen and co-workers 1934, 1935, Brück 1936, Goche and Wilman 1939, Goswami 1954), and partial disorientation about certain axial directions of epitaxially grown zinc oxide on zinc sulphide (Evans and Wilman 1950), cadmium on copper (110) (Goswami 1950), silver on sodium chloride (Brú and Gharpurey 1951) are, no doubt, examples of such strain relief.

The effect of interfacial stress could be expected to extend not only to the overgrowths but also to the substrate surface and below this. Hitherto, however, this latter effect has not been observed.

Recently we have found that during the anodic treatment of a single crystal of copper in an acidified halide solution, considerable disorientation of the copper surface was caused. Electropolished and etched (111) and (110) faces of copper, which were free from strain and had a high degree of lattice perfection, yielded electron diffraction patterns (figures 1 and 2, Plate) after anodic treatment in a solution of 10 g l.$^{-1}$ potassium bromide and 25 g l.$^{-1}$ sulphuric acid at current density about 40 mA cm^{-2} or more for a second or two at room temperature.

The surface was partially covered with thin films. Extensive Cu (111) twinning spots as well as continuous rings due to copper (confirmed by comparison with graphite $11\bar{2}0$) were observed and are evidently caused by the disorientation of the substrate surface. Some arcs due to the (111) orientation of cuprous bromide and cuprous oxide are also present in the pattern. On the Cu (110) face a similar disorientation of the surface crystal structure was also observed, although neither cuprous bromide nor cuprous oxide was detected in the pattern (figure 2). This may possibly be due to the removal of coating during washing of the specimen after preparation. Similar disorientation of Cu (110) face was observed during anodic treatment in concentrated hydrochloric acid solution (Goswami 1950).

The above experiment was repeated, after much electropolishing and etching (to remove any possible surface strain). The results were similar though the pattern of cuprous bromide was more prominent and no cuprous oxide was observed. Cuprous bromide grew epitaxially with (111) orientation but its axis was rotated through 30° with respect to the copper (figures 3 and 4). The deposits also formed extensive $\{111\}$ twinned crystals and this is confirmed from the pattern when the beam was along $[2\bar{1}\bar{1}]$ aximuth of copper (figure 4).

The above disorientation of the copper surface cannot be explained as due to the reduction of cuprous bromide or cuprous oxide, as the treatment was anodic and no reduction was possible. The only cause, as it appears to us, was the interfacial stress, which ultimately relieved itself by the disorientation of the single crystal surface or by the formation of twinned crystallites.

REFERENCES

BRÜ, L., and GHARPUREY, M. K., 1951, *Proc. Phys. Soc.* A, **54,** 283.
BRÜCK, L., 1936, *Ann. Phys., Lpz.*, **26,** 233.
COCHRANE, W., 1936, *Proc. Phys. Soc.*, **48,** 723.
EVANS, D. M., and WILMAN, H., 1950, *Proc. Phys. Soc.* A, **63,** 298.
FINCH, G. I., WILMAN, H., and YANG, L., 1947, *Disc. Faraday Soc.*, **1,** 144.
GOCHE, O., and WILMAN, H., 1939, *Proc. Phys. Soc.*, **51,** 625.
GOSWAMI, A., 1950, *Ph.D. Thesis*, University of London, 1954, *J. Sci. Industr. Res.* B, **13,** 677.
KIRCHNER, F., and LASSEN, H., 1935, *Ann. Phys., Lpz.*, **24,** 173.
LASSEN, H., 1934, *Phys. Z.*, **35,** 172.
LASSEN, H., and BRÜCK, L., 1935, *Ann. Phys., Lpz.*, **22,** 65.
WILMAN, H., 1940, *Proc. Phys. Soc.*, **52,** 323.

REVIEWS OF BOOKS

Molecular Beams, by K. F. Smith. Pp. x + 133. (London: Methuen's Monographs on Physical Subjects.) 8*s*. 6*d*.

Dr. Smith's monograph replaces the earlier volume in this series by Dr. Fraser, published in 1937. Few subjects can have developed so rapidly in the interval. The introduction of radio-frequency resonance methods, by which very small multiplet intervals are measured directly, has transformed the technique, and has enabled measurements of nuclear magnetic moments to be made with an extraordinary precision.

The book covers a very wide field of work, from the use of molecular beams as sources for optical spectroscopy to the most refined of the resonance experiments. The author is to be congratulated on dealing with all this material without losing touch with the needs of students, for whom this series is primarily intended. The long section on magnetic deflection, with its detailed consideration of effective magnetic moments in low and high fields, is especially useful in clearing up many points that students find difficult. The book is well produced, and with 133 pages, including 37 figures and 12 tables, is commendably cheap.

A. H. COOKE.

Classical Electricity and Magnetism, by W. K. H. Panofsky and M. Phillips. Pp. xi + 400. (Cambridge (Mass.): Addison-Wesley, 1955.) $8.50

The student of electromagnetism has at his disposal textbooks of rare excellence: Becker's revision of the old classic 'Abraham–Föppl', Sommerfeld's inspiring lectures and, in a more formalistic vein, the recent treatise by Landau and Lifshitz. In a domain so thoroughly and authoritatively worked out it would be hard to innovate. Indeed, one wonders whether there is any real need for yet another textbook of comparable scope.

The book under review appears to respond to some specific requirement of the American University curriculum, and makes no ambitious claims to originality. It follows in an eclectic way the well-established lines, with judicious emphasis on the more modern aspects. Besides the fundamental laws comprehended in Maxwell's equations, it covers the special theory of relativity, the theory of waveguides, the radiation field of accelerated charges, the elementary theory of scattering and dispersion, and the treatment of electromagnetic interactions from the Hamiltonian point of view. It thus brings together in a clear, logical disposition a most felicitous selection of topics, which is perhaps not found so nicely balanced in the other treatises. An excellent feature is the list, at the end of each chapter, of books recommended for further reading, with brief indications of what is to be looked for in them. The examples are made interesting by being often taken from 'actual life', i.e. from problems met with in the discussion of experiments in atomic and nuclear physics.

The treatment is throughout clear and concise, but surprisingly abstract, in the sense that the physical aspect of the problems is only just evoked in laconic sentences rather than properly explained.† It seems that the authors have in mind an audience who has already been taught the basic facts without mathematics, and has only to be shown how to translate them into the mathematical language. From the latter point of view, they again presuppose a rather solid knowledge of vector analysis in the Gibbsian notation, which is used from the beginning without explanation. The book is therefore not self-contained, but for the student who has the required physical and mathematical preparation, it may be recommended as a reliable and practical introduction. It has good diagrams and the printing of text and formulae is of exquisite neatness.

---

† For some reason, the authors make an exception in favour of Michelson's experiment, of which they give a detailed and interesting account.

L. ROSENFELD.

Principles of Colour Photography, by R. M. EVANS, W. T. HANSON, Jr., and W. L. BREWER. Pp. xi + 709. (New York: John Wiley; London: Chapman and Hall, 1953.) 88*s*.

So complex is the modern outworking of the simple concept of colour photography, that a book of 709 pages is needed to discuss its principles, and the conscientious reader finds his mental powers severely taxed, as he passes from physics to psycho-physics, from psycho-physics to chemistry, from chemistry back to physics, and on to mathematics, psychology, and aesthetics, without ever straying from his chosen realm of colour photography. And this surely is one of the fascinations of the subject, its endless variety.

The first four chapters of this book deal with the response of the eye to colour, the specification of colour, and the bearing which these concepts have on the fundamentals of colour photography. The next two chapters deal with some of the fundamentals of photography such as latent image formation and densitometry. There follows a chapter in which the different methods of forming *coloured* images are described, and this leads to a review, in the next chapter, of colour photographic systems and a short history of the subject. The next four chapters discuss types of colorants, optical characteristics of colorants, colour densitometry, and colour sensitometry. That this takes 489 pages will bear witness to the thoroughness with which the foundations of the subject have been laid. But one criticism which the present reviewer has to make is that so much is included that the wood sometimes cannot be seen for the trees, and it might have been preferable to have omitted some of the material which is already available in other standard works.

In chapters 13 and 14, however, results of great interest and real value to any worker in this field are presented. The effects of varying the spectral sensitivities, the colorants, and the characteristic curves of colour photographic systems are carefully analysed and discussed, and one feels that the writers have come successfully to grips with one of the most intractable problems in the field.

The last four chapters, on duplicating, copying, and reproduction theory for additive and subtractive systems, tend to be more exercises in matrix algebra than guidance for practice, and in particular it seems doubtful if the emphasis on systems having linear characteristic curves and six masks is warranted.

In a work that is otherwise so complete, it is surprising to find no mention of highlight masking (an important technique in copying and duplicating). The very considerable variability of colours in original scenes is not mentioned, nor is the related problem of the permissible tolerances in colour reproduction given more than passing reference.

These are, however, but minor criticisms of a work which will undoubtedly take its place alongside Mees' *Theory of the Photographic Process* as a standard reference book. The authors are to be congratulated on the thoroughness, accuracy and clarity which they have achieved throughout this excellent production.
R. W. G. HUNT.

Optique électronique I: *Lentilles électroniques*, by P. GRIVET, M. Y. BERNARD and A. SEPTIER. Pp. 184. (Paris: Bordas, 1955.) *fr*. 1450.

Professor Grivet is setting out to produce a complete treatise on the theory and practice of electron optics in three volumes, of which this is the first. Succeeding volumes will deal with microscopes, spectrographs and accelerators for charged particles. The present text provides an excellent introduction to the theory, properties and design of electrostatic and magnetic electron lenses. The level of treatment is postgraduate rather than undergraduate, assuming our acquaintance with vector analysis. It is a research monograph, which successfully avoids being a technical handbook whilst providing a great deal of practical detail to illustrate and supplement the theoretical exposition.

Chapters on electrostatic and magnetic fields include methods of accurate measurement as well as mathematical evaluation. The properties of electrostatic and magnetic lenses are described in the next two chapters, including the relativistic treatment. A full treatment of lens aberrations is preceded by an illuminating discussion of the similarities and differences between glass and electron optics. The design and construction of electrostatic and magnetic lenses are then treated in some detail. The final chapter deals with the recently developed method of ' strong focusing ' and its value in high energy particle accelerators. The text is clear and concise throughout. The mathematics is well displayed and line diagrams are freely used (108 in all). By and large it is the best short account of electron lenses that has yet appeared. If a fault can be found, it is that magnetic lenses do not receive the proportion of attention corresponding to their many practical applications, whereas electrostatic lenses are very fully treated; but this is an understandable consequence of the main preoccupation of the authors' laboratory. It must be said that they have done their best to be fair to the magnetic type. The book contains a short index and a good selective bibliogaphy. To English eyes, a book which is liable to get as much use as this would be better bound in a stiff rather than a paper cover.
V. E. COSSLETT.

Optical Workshop Principles, by C. DÉVÉ, English translation by T. L. TIPPELL, 2nd Edn. Pp. xxiv + 436. (London: Hilger and Watts, Ltd., 1954.) 42*s*.

The late Charles Dévé has made a major contribution to optical workshop techniques. In this book he records in a concise manner his own experiences of optical working processes. This is the second English edition of a well-known book and it contains much new material. The translation has been revised and the layout and typography improved.

In Part I glass types, abrasives, cements and allied materials are considered. A chapter is devoted to ' surfacing ' in order to lay the foundation for the main discussion on surfacing in Part II. There is no hesitation to discuss the more difficult optical problems, for example the surfacing of small deep curves and cylindrical surfaces.

This is the only book dealing with a coherent theory of surfacing and much of the work Dévé describes is derived from the five fundamental theorems: these are set out in Part II " for the use of works managers and senior workmen ". The application of these theorems has enabled Dévé to devise regimes of uniform wear for the production of all types of surfaces. Chapter VI deals with optical tests in the workshop; and it is unfortunate that in this edition a description of the lens and prism interferometer was omitted. This type of interferometer is now regarded as a standard piece of test gear in a modern workshop for precision optics. The use of test plates is treated in great detail.

In Appendix B some mechanisms are described for the making of non-spherical surfaces. The value of the appendix in an important branch of optical working is doubtful because the mechanisms " have not yet been completed or even constructed ". The geometrical properties of some conic sections are considered in order to establish methods for the production of form tools, which can be used in a lathe to make smoothing tools for glass surfaces. It is also suggested that ' hard metal ' tipped form tools could be used for turning glass.

Scouring (Caillebottage) is advocated as a method for trueing non-spherical surfaces, and to this end an ' Epicyclic Oscillating Mechanism ' has been devised to surmount the difficulty of realizing mechanically the conditions which ensure wearing away of privileged zones in a smooth manner.

A surprising but welcome addition in a book on ' Workshop Principles ' is a twenty page appendix on ' The Physical Nature of Light ' written by Charles Guadet. This is a factual and concise introduction of importance to the optical worker in order that he may understand and appreciate the fundamentals upon which interference tests are based. In the compass of a few pages will be found the essential information, which may take many hours to glean from a textbook on Physical Optics.

The book of 436 pages is well presented and well bound: it contains a few printing errors. The list of contents is brief and could be made more comprehensive. The index of this volume is adequate. This book will be studied and consulted over and over again by optical workers and should find its place in the library of many individuals and optical institutions. K. W. BRITTAN.

High Vacuum Technique (3rd Edn Revised), by J. YARWOOD. Pp. viii + 208. (London: Chapman and Hall, 1955.) 25*s*.

The applications of high vacuum form but a small part of this book, since they are now both numerous and exacting in specialist knowledge. Instead, Mr. Yarwood has concentrated on the production and measurement of high vacua, the measuring of pumping speeds, and the properties of materials used in vacuum work. This concentration is to be applauded, when the result is such a helpful compilation as Mr. Yarwood has achieved. There are several up-to-date tables of such data as the performances of the many pumps now commercially available: it is very useful to have these data assembled in one volume. Various recent aspects of high vacuum work such as gas-ballast pumps, pump fluids, valves and seals, and leak detectors, are well treated. The references and the diagrams are both plentiful and good.

This is a book, not an expensive one, which will be very helpful to any worker, from the research student stage onwards, who has to use high vacua.

R. V. JONES.

Statistical Mechanics of Irreversible Change, by R. T. Cox. Pp. vii + 130. (Baltimore: John Hopkins Press; London: Cumberlege, 1955.) 40*s*.

In this monograph the author summarizes his views which have partly been published in previous articles. He aims at formulating a general theory governing the rate of approach to thermal equilibrium. The thermodynamics of irreversible processes is not suitable as a basis for his approach since its scope is not wide enough; nor does the author subscribe to a fully microscopic theory which would, as he believes, be encumbered by irrelevant detail. " Between these two methods ", the author continues, " there is a third which ignores the molecular details of the structure but takes account of the microscopic details of change. This is the statistical method of Gibbs in which the entire thermodynamical system is taken as the statistical unit and the structural ensemble is a collection of like systems. It is this method which will be followed here."

In accordance with this programme the author considers probability distributions for the quantum states of macroscopic systems and the probabilities for transitions between these states. It is asserted that the transition probabilities comply with the principle of microscopic reversibility. From these premisses the theory of a wide class of phenomena is readily deduced, including the Brownian movement—where the author has particularly interesting points to make—the general rate of approach to equilibrium in the linear approximation, Onsager's symmetry rules, thermal conduction, thermomechanical effects and other important types of irreversible processes; the conventional equilibrium theory is recovered as a by-product.

The success of the author's approach is, however, qualified by doubts regarding the validity of his premisses. They are, strictly speaking, refuted beforehand by Gibbs (*Elementary Principles in Statistical Mechanics* (New York 1902), Ch. XII; also R. C. Tolman, *Principles of Statistical Mechanics* (Oxford 1950) § 51): the change in time of Gibbs' ensembles is necessarily reversible; ensembles are suitable for specifiying irreversible change only when ' coarse grained ' distribution functions are considered. In fact, the principle of microscopic reversibility has been established only for relatively small subsystems of a macroscopic system; its application to the macroscopic system as a whole would be difficult to justify and author's deductions in sections 3, 8 and 15 cannot but be regarded as dubious. The major part of the book under review stands, however, on safer ground since the arguments remain valid if the principle of microscopic reversibility is—legitimately—applied to subsystems only, which may of course consist of a large number of molecules.

The author's attempt at unifying the theory of irreversibility deserves serious attention in spite of these critical remarks. It is likely that much of his present work will persist even when the foundations of the theory are better secured. This monograph will hardly appeal to the general reader but should be useful for research workers in related fields. Layout, print and binding of the book are gratifying. R. EISENSCHITZ.

Microstructures of Diamond Surfaces, by S. TOLANSKY. Pp. viii + 67. (London: N. A. G. Press, 1955.) 40*s*.

The last ten years or so have seen the development and application, by Tolansky and his school, of improved methods of great power, precision and beauty for examining the micro-relief of surfaces. Here these methods are employed to document the growth patterns, and other crystallographically significant features of diamond, with much instructive detail, including new discoveries and much more exact information about features already known. To the reviewer, who finds the interpretation of the growth of diamonds a particularly fascinating problem, it is a slightly invidious task, on behalf of his fellow Fellows of the Physical Society, to look this gift horse in the mouth.

The text accompanying these photographs has been written with a wide variety of readers in mind. In the author's especial field of methods of microsurvey he achieves a high standard of elementary exposition; but the need for simplicity does not excuse such statements as " 20 Å is the size of a mica molecule" (p. 6), nor is this much improved by expansion to " a single molecular lattice of mica " (p. 14); and in matters of crystal theory the author from time to time hits the nail squarely on the thumb. His interpretative deductions within this field, right or wrong, are rarely logically solid. The growth patterns of diamond on its octahedral faces are so dominated by the peculiar 'trigons' that the way in which growth layers commence has so far evaded detection. To reject a spiral commencement by saying that *on the contrary* diamond grows by spreading layers (or 'plane waves' as Tolansky puts it—p. 30) gets us nowhere. The etch habit (p. 40) cannot be (221), for that would put the etch pits in the wrong orientation. The markings (plate 55) which characteristically meet without crossing, cannot be normal slip-lines, as suggested : all the facts about them are compatible with their being twin lamellae, and the one objection raised against this has no force. Those (plate 76) which cross, cannot be the traces of growth surfaces, as suggested, but may well be slip-lines. Plate 83, and others, show excellent examples of growth-surface traces, correctly identified.

Much more important is the question whether the objective documentation is thoroughly reliable. Unfortunately, this is not invariably the case. It is inconceivable that plates 63, 64, 65 show the state of a crystal after successive hours of etching : 64 is 63, slightly out of focus. Plate 53 is almost certainly a double exposure, slightly shifted. Plates 59 and 60 certainly do not show exact correspondence, as stated on page 37 : one of them has probably been accidentally replaced by a similar example. These are comparatively minor failings in a valuable record to which all future workers on diamond must pay attention.

To those who use diamond as a tool, the information in the last two chapters, on 'ring-crack' damage to diamonds, and on the surface condition of mechanically polished diamonds, should be specially worthy of study.

F. C. FRANK.

Le Microscope à Contraste de Phase et le Microscope Interférentiel, by M. FRANÇON. Pp. v + 149. (Paris : Centre National de la Recherche Scientifique, 1954.) *fr.* 1,000.

This little book is divided into two major parts. The first of these, which is not suggested in the title, is concerned with the mode of image formation in the conventional microscope. Starting with the case of a luminous point object, the effects on the image of diffraction and of instrumental imperfections are considered, together with such topics as apodization and the consequences of central obstruction. This is presented in descriptive manner without mathematical derivations, results only being quoted. A second chapter deals with the images of extended objects in coherent and incoherent illumination, limits of resolution and perception, and the influence of defects of the eye itself.

The second part, itself in two sections, covers the field of the observation of transparent objects. The principles of the phase contrast and interference microscopes are set out, together with descriptions of various forms of both these instruments. Some applications of the phase contrast microscope are mentioned, together with illustrative photographs and brief descriptions of the special techniques involved.

An extensive bibliography is appended, but unfortunately there is no index.

The first part renders this book valuable to all users of the microscope who are not optical specialists. The effects of the various aberrations are described with typical Gallic clarity and several photographs, and the difference between the resultant interference pattern and the geometrical aberration figure is well brought out. That the discussion is hardly the 'theoretical study' promised in the Introduction is scarcely a disadvantage in a book of this type.

It is regrettable, however, that similar attention is not given to the structure of phase contrast and interference images. The nature of the 'halo effect' is not very clearly brought out, and Figure 144 (*c*), which purports to show this, is incorrect. This omission could largely have been repaired by further photographs illustrating the same objects by phase contrast with linear and annular phase plates, and by interference contrast. This could have been done without great extra cost by omitting the one coloured plate illustrating interference contrast, which is not particularly informative and is not even referred to in the text.

The standard of the figures in this book, although adequate for clarity, will disappoint those who know that Professor Françon can do as well freehand on the blackboard.

Notwithstanding the above criticisms, this book can be recommended for the reading of the practical microscopist as a contribution to his scientific education, rather than as a laboratory handbook. J. DYSON.

CORRIGENDUM

Luminescence and Electrical Conductivity associated with Defects in Alkali Halide Crystals, by J. EWLES and J. C. STEAD (*Proc. Phys. Soc.* B, 1956, **69,** 396).

In this paper it was stated : "In previous work a.c. methods have been limited by stray capacitance effects to resistances of not more than about 10^7 ohm so that investigations could only be carried out with alkali halides either at temperatures above about 300°c or with the unsatisfactory d.c. method."

It has now been pointed out to the authors that Macdonald (1955) has described the use of an a.c. method for measuring the resistances of alkali halides up to at least 10^{12} ohm.

MACDONALD, J. R., 1955, *J. Chem. Phys.*, **23,** 275.

CONTENTS OF SECTION A

PAGE

Dr. C. B. O. Mohr and Mr. B. A. Robson. Calculations with the Optical Model of the Nucleus 365

Dr. D. H. Whiffen. Intensities in the Raman Spectrum of Benzene. . . . 375

Mr. S. Messelt and Dr. A. Storruste. Elastic Scattering of Gamma Rays . 381

Mr. R. S. Leigh. The Augmented Plane Wave and Related Methods for Crystal Eigenvalue Problems 388

Dr. C. R. Worthington and Dr. S. G. Tomlin. The Intensity of Emission of Characteristic X-Radiation 401

Research Notes:

Mr. K. P. Nicholson. Time Delays in Boron Trifluoride Proportional Counters 413

Dr. B. Andersson. Precision Wavelength and Energy Measurements of Gamma Rays from ^{152}Eu, ^{154}Eu, ^{153}Sm and ^{165}Dy 415

Dr. A. W. Aikin. Neutron Production by Electrons in Uranium and Uranium–Beryllium 419

Dr. B. H. Bransden and Mr. J. S. C. McKee. The 1s–2s Excitation of Hydrogen by Electron Impact 422

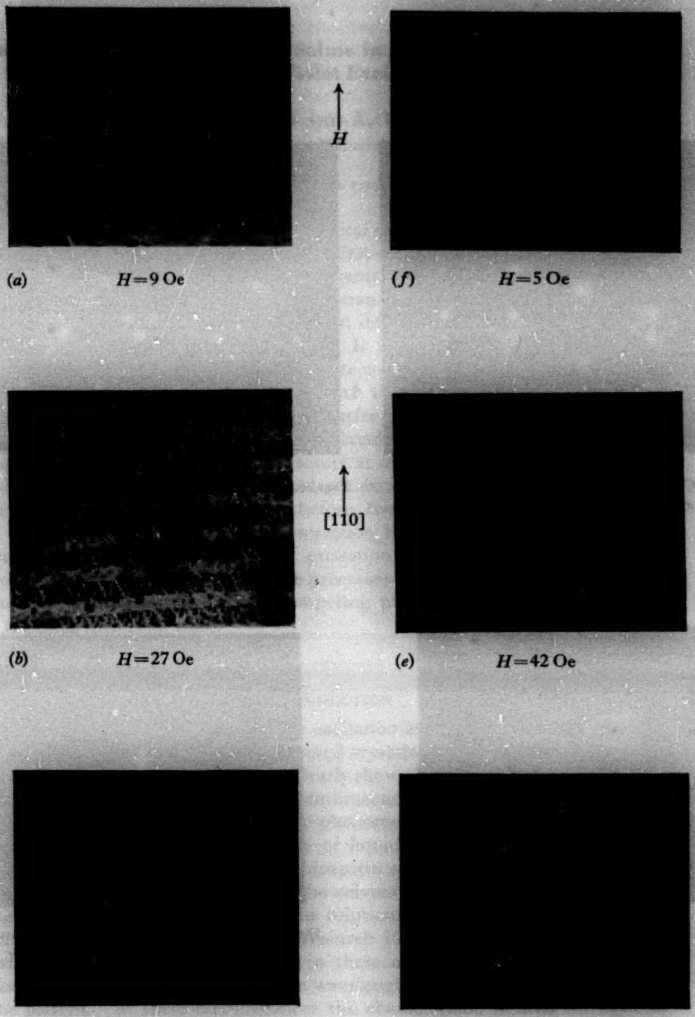
Letters to the Editor:

Prof. S. J. Khambátá. Raman Spectrum of Magnesium Oxide 426

Dr. E. P. George. On Star Production by μ-Mesons Underground . . 428

Reviews of Books 430

Contents of Section B 436



(*a*) $H=9$ Oe

(*f*) $H=5$ Oe

(*b*) $H=27$ Oe

(*e*) $H=42$ Oe

(*c*) $H=255$ Oe

(*d*) $H=257$ Oe

Figure 2. Bitter figures on (110) surface; (*a*)–(*c*) field increasing, (*d*)–(*f*) field decreasing (×112)

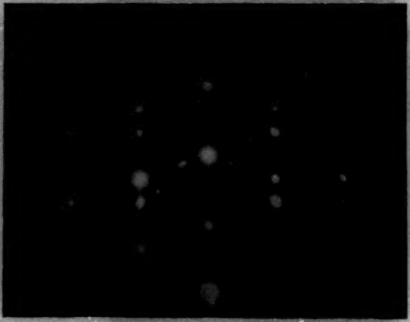


Figure 1. Cu (111) face anodically treated; beam along [1$\bar{1}$0] direction of copper.

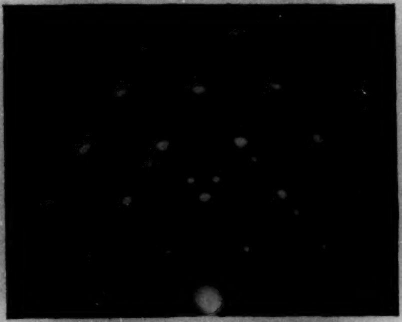


Figure 2. Cu (110) face anodically treated; beam along [1$\bar{1}$0] direction of copper.

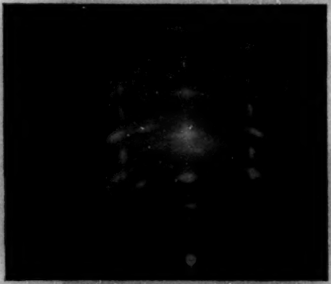


Figure 3. Cu (111) face anodically treated; beam along [1$\bar{1}$0] direction of copper.

Figure 4. Cu (111) face anodically treated; beam along [2$\bar{1}\bar{1}$] direction of copper.

Energy Transfer from Solvent to Solute in Liquid Organic Solutions under Ultra-Violet Excitation

By S. G. COHEN and A. WEINREB

Department of Physics, Hebrew University, Jerusalem, Israel

MS. received 19*th September* 1955, *and in revised form* 2*nd January* 1956

Abstract. The transfer of molecular excitation energy from solvent to solute in liquid organic solutions under ultra-violet excitation has been studied as a function of different parameters such as solute concentration, exciting wavelength and dilution of absorbing solvent with a second transparent solvent. By measuring the fluorescent efficiencies of the solutions at different exciting wavelengths the intrinsic transfer efficiency can be found. It is found that the transfer probability is accurately proportional to the solute concentration, and is independent of exciting wavelength in the region in which the solvent absorbs appreciably. An investigation of the quenching of the transfer process by carbon tetrachloride shows that an average time of about 10^{-9} second must elapse before the solvent excitation energy is transferred to the solute at optimum solute concentrations. It is shown that in diluted systems the solvent excitation energy is not exchanged between different solvent molecules before transfer. In undiluted systems there is evidence for a short range interaction between the solvent molecules leading to a relatively weak transfer of excitation energy through neighbouring solvent molecules. It is shown that the processes of natural solvent fluorescence and solvent–solute transfer are not competing processes which can take place from the same molecular state.

---

§ 1. Introduction

Examples of transfer of molecular excitation energy between molecules of different kinds in solid organic mixed crystals have been known for some time. For example, this is very clearly shown by the fluorescent behaviour of a solid solution of naphthacene in anthracene under ultra-violet excitation (Bowen 1945). Renewed interest in the phenomena of energy transfer between molecules was created by the discovery of liquid organic scintillating systems (Kallmann and Furst 1950, Ageno, Chiozzotto and Querzoli 1950, Reynolds, Harrison and Salvini 1950), in which the solvent molecules are excited by the passage of nuclear radiations through the solutions.

In a preliminary note (Cohen and Weinreb 1954), the authors reported the existence of a similar energy transfer in these organic liquid solutions when excited by ultra-violet radiation at those wavelengths which are mostly absorbed by the solvent. In these circumstances the characteristic fluorescent light of the solute molecules is emitted and a high quantum fluorescence yield may be obtained. Because of the close resemblance in the general behaviour of these organic solutions under both kinds of excitation, we are led to believe that the actual transfer process is identical in both cases. The initial processes involved in the case of excitation by ionizing radiation are of course very complicated.

In the case of ultra-violet excitation, however, the primary process, namely the straightforward excitation of the solvent molecules, is much simpler. Therefore one might expect that a detailed investigation of the behaviour of liquid systems under ultra-violet excitation might lead to a better understanding of the various processes involved, including the nature of the energy transfer.

The fraction of excited solvent molecules which transfer their energy to the solute molecules shall be called the intrinsic transfer efficiency and shall be denoted in this paper by ϵ. Measurements of the fluorescent intensities under excitation by ionizing radiation or by ultra-violet radiation which is absorbed mostly by the solvent, do not permit a determination of this important quantity ϵ, since the fluorescent intensity is determined both by ϵ and by the degree of solute quenching, which may be produced either by molecules of the solvent ('external quenching') or by interaction with other solute molecules ('self-quenching'). Nevertheless, by making measurements of the fluorescent efficiency of the solution, both at exciting wavelengths corresponding to almost complete absorption by the solvent and also at exciting wavelengths corresponding to almost complete absorption by the solute, a separation of the various factors can be made, and ϵ can be determined.

The main purpose of this investigation was to measure the intrinsic transfer efficiency as a function of various variables such as exciting wavelength, concentration, presence of quenching agents, in simple and complex liquid systems, in order to learn as much as possible about the nature of the transfer process, about which little is known with any certainty.

Since our preliminary note, the works of Furst and Kallmann (1954, 1955) have appeared, in which the usefulness of the experimental approach using ultra-violet excitation has been illustrated.

§ 2. EXPERIMENTAL METHOD

2.1. *Principles*

The general experimental method adopted in separating out the various factors which influence the fluorescence efficiency under ultra-violet excitation can perhaps best be illustrated by considering a particular system, for example, a solution of p-terphenyl in toluene at a concentration of 1 g l.$^{-1}$. Such a system is known as a fairly efficient scintillator. If this solution is irradiated with ultra-violet radiation at wavelength of about 3000 Å, effectively all the absorbed quanta are absorbed by the solute alone. Hence no solvent–solute transfer effects are involved, and the fluorescence efficiency of the solution at this exciting wavelength will be determined by the external and self-quenching of the solute fluorescence. On the other hand, at an exciting wavelength of about 2500 Å, 97% of the incident quanta are absorbed by the solvent and only 3% by the solute. The high solute fluorescence efficiency obtained at this wavelength (Cohen and Weinreb 1954) therefore points to an efficient solvent–solute transfer process. Under the assumption that the fluorescent efficiency of an excited solute molecule is independent of its energy of excitation between 2500 and 3000 Å, the ratio of overall fluorescence efficiency obtained at 2500 Å (after corrections for the small solute absorption) to that obtained at 3000 Å gives directly the intrinsic transfer efficiency. It is, of course, impossible to verify this assumption directly for terphenyl in toluene, but the fluorescence efficiency of terphenyl can be examined as a function of wavelength when dissolved in a solvent which does not absorb throughout the range of wavelengths. This was investigated using chloroform and cyclohexane as solvents which do not absorb significantly in this range, and no variation was found between 2500 and 3000 Å. Chloroform quenches the solute fluorescence by about 35% relative to toluene, whereas in cyclohexane the external quenching is less than in toluene. Hence it is reasonable to assume that the fluorescent efficiency of an excited molecule of terphenyl is independent of the energy of excitation within the range considered.

In general, if I_1 and I_2 are the solute fluorescent intensities obtained at exciting wavelengths λ_1 and λ_2 corresponding to absorption by the solvent and by the solute respectively, η is the quantum yield of the solute (assumed to be independent of wavelength) and n_1 and n_2 are the flux of incident quanta at these two wavelengths, then $I_1 = A\eta n_1 \epsilon$, $I_2 = A\eta n_2$ where A is a constant of the apparatus and independent of wavelength and ϵ is the intrinsic transfer efficiency as given above. Then

$$\epsilon = \frac{I_1/n_1}{I_2/n_2}. \qquad \text{......(1)}$$

Since in the systems under consideration, transfer of energy takes place between solvent molecules whose first excited states are higher than those of the solute molecules, at sufficiently long wavelengths the absorption of the solvent will be negligible compared with that of the solute. Hence in these systems a measurement of ϵ by the above method can, in principle, always be made.

2.2. *Experimental Procedure*

All the measurements of fluorescent intensity were carried out using a Beckman quartz spectrophotometer, model D.U., with minor modifications. This instrument was used both in order to monochromatize the incident radiation and also to measure the intensity of the fluorescent radiation. Measurements were made either using the No. 2260 Beckman mercury arc lamp as a source, or, when higher intensities were necessary, a Hanovia U.V.S. 250 high pressure mercury lamp. For convenience, the fluorescence was measured in the same direction as the incident radiation. In most cases the thickness of solution used (one centimetre) was sufficient to absorb practically all the exciting radiation, even at the longer wavelengths where the solvent does not absorb. Nevertheless, appropriate filters were always used in order to eliminate completely any direct contribution from the primary exciting radiation. The liquids were contained in rectangular quartz cells. Corrections, which were generally small, were carried out for the following effects when necessary: (i) background intensities due to scattered light from the monochromator: this is a function of the monochromator setting and slit width; (ii) the relatively small amount of fluorescence excited in the entrance window of the quartz cells, at exciting wavelengths shorter than 2600 Å; (iii) geometrical and self-absorption effects arising from the variation in absorption coefficient among the various solutions.

In addition to comparing I_1 and I_2 it is necessary to compare the ratio of numbers of incident quanta n_1 to n_2 at the different wavelengths, in order to obtain ϵ in accordance with equation (1). For this purpose the intensities of the incident primary radiation were monitored by measuring the fluorescence induced in an aqueous solution of fluorescein. According to the measurements of Vavilov (1922) and more recently of Fuchslocher and Glaser (1954), the fluorescence yield of fluorescein is independent of exciting wavelength in the range under consideration.

In the experiments to be described, p-terphenyl or anthracene were used as fluorescent solutes, toluene and anisole as solvents absorbing the primary radiation and transferring to the solute, and cyclohexane as a solvent transparent even for the short wavelengths. Similar qualitative behaviour, however, was in general exhibited using liquid systems made up of other common organic solvents (e.g. benzene, xylene, as absorbing solvents and chloroform, hexane, alcohol and acetic acid as non-absorbing solvents). Thus the authors believe the results to have a general validity, although particular systems were chosen for detailed investigation in order to demonstrate most clearly and conveniently particular phenomena.

§ 3. EXPERIMENTAL RESULTS

3.1. *The Influence of Concentration on Transfer Efficiency*

These experiments were carried out using two systems, a solution of p-terphenyl in toluene, and a solution of anthracene in anisole. It will be seen that although these solutions possess fluorescence efficiencies which differ almost by an order of magnitude for equivalent solute concentrations, the form of the dependence of transfer efficiency on concentration and also the absolute values of this quantity are similar.

Figure 1 (*a*) shows the relative fluorescence efficiency as a function of the concentration of p-terphenyl in toluene for exciting wavelengths of 2537 Å (curve 1) and 3130 Å (curve 2) respectively. It follows from the discussion of the experimental methods (§ 2.1) that the ratio between the ordinates of the two curves at a given concentration yields directly the intrinsic efficiency at this concentration. This is shown in figure 1 (*b*). As to be expected the transfer efficiency increases with increasing concentration reaching the high value of about 88% at a concentration of $5\,\mathrm{g\,l.^{-1}}$, i.e. at this concentration very nearly all the excited solvent molecules transfer energy to the solute. Curve 1 in figure 1 (*a*), which represents the overall fluorescent efficiency after solvent absorption (and which one might, *a priori*, expect to resemble the variation of scintillation efficiency with concentration under gamma-ray excitation) does not differ very much from curve 1 (*b*) showing the behaviour of ϵ, since curve 2 in figure 1 (*a*) shows that the quenching of terphenyl in toluene only changes slightly with concentration, i.e. the self-quenching must be very small.

Figures 1 (*c*) and 1 (*d*) show the corresponding curves for anthracene in anisole for exciting wavelengths 2650 Å and 3340 Å respectively. Curves 1 and 2 in figure 1 (*c*) are now quite different from the corresponding curves in figure 1 (*a*). In particular, the fluorescent efficiency of anthracene is low, and also strongly concentration dependent as shown by curve 2 of figure 1 (*c*). On the other hand, the transfer efficiency curve in figure 1 (*d*) shows the same general behaviour as in the previous example, and the transfer efficiencies at high concentrations have similar values (88% and 70%). Owing to the self-quenching of anthracene which is very concentration dependent (curve 2 in figure 1 (*c*)) the overall fluorescence efficiency, in this case (curve 1, figure 1 (*c*), reaches a maximum and then decreases at sufficiently high concentrations.

Curves 1 of figures 1 (*a*) and (*c*) do indeed bear a very close resemblance to the curves obtained by other workers showing the dependence of scintillation

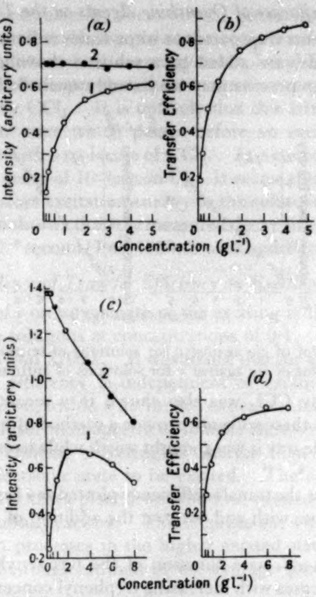


Figure 1. (*a*) Variation of fluorescence intensity of *p*-terphenyl in toluene with concentration of *p*-terphenyl; 1, 2537 Å; 2, 3130 Å.
(*b*) Variation of energy transfer efficiency in solutions of *p*-terphenyl in toluene with concentration of *p*-terphenyl.
(*c*) Variation of fluorescence intensity of anthracene in anisole with concentration of anthracene; 1, 2650 Å; 2, 3340 Å.
(*d*) Variation energy of transfer efficiency in solutions of anthracene in anisole with concentration of anthracene.

efficiency with solute concentration. (While no data is available for anthracene in anisole, the curve for anthracene in xylene (Kallmann and Furst 1950) shows the same shape as curve 1, figure 1 (*c*). This is a strong justification for assuming that the final transfer process after gamma-ray excitation is identical with that taking place after ultra-violet excitation.

If the probability of transfer is proportional to the concentration c of the solute then c/ϵ should be a linear function of c, since ϵ is then given by $\epsilon = ck/(ck+q)$, where ck is the probability of transfer and q is the probability of loss of excitation energy of the solvent molecules by other means. One concludes from figures 2 (*a*) and (*b*), which show the linear plots, that the transfer probability is accurately proportional to the concentration for both the solutions investigated. Earlier attempts (Kallmann and Furst 1950) to demonstrate this relation in the case of gamma-ray excitation were complicated by the concentration dependence of quenching of solute fluorescence described above and more complicated equations with a greater number of parameters have to be introduced.

3.2. *Influence of Quenching Agents on the Transfer*

When certain solvents transparent to ultra-violet radiation, such as chloroform or carbon tetrachloride, are added to a solution showing fairly good transfer properties, the transfer process may be strongly quenched. The quenching of the transfer process by CCl_4 was also shown in a recent paper by Furst and Kallmann (1955). As these solvents provide a method of influencing the transfer process in a controllable way it was thought worth while to examine this quenching quantitatively.

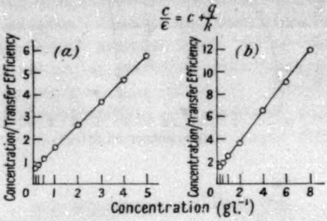


Figure 2. (*a*) Plot of c/ϵ against c for solutions of terphenyl in toluene.
(*b*) Plot of c/ϵ against c for solutions of anthracene in anisole.

In figure 3 (*a*) we see the transfer efficiency plotted as a function of the terphenyl concentration in toluene with and without the addition of 3 g l.$^{-1}$ CCl_4 (ϵ_q and ϵ respectively).

Figure 3 (*b*) shows ϵ_q/ϵ as a function of the terphenyl concentration. The influence of CCl_4 decreases with increasing terphenyl concentration. Figure 3 (*c*) shows a plot of c/ϵ_q as a function of c, the terphenyl concentration. The linear relation shows that the equation of the former section (with a new value of the quenching coefficient $\bar{q}$) can also be used to describe the competition between the solvent–solute transfer and between the quenching processes which in this case includes the collision quenching of the solvent by molecules of CCl_4.

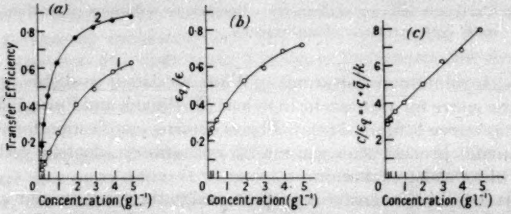


Figure 3. (*a*) Variation of transfer efficiency in solutions of *p*-terphenyl in toluene with concentration of *p*-terphenyl; 1, with addition of 3 g l.$^{-1}$ CCl_4; 2, without CCl_4.
(*b*) Variation of ϵ_q/ϵ with concentration of *p*-terphenyl.
(*c*) Plot of c/ϵ_q against c for solutions of *p*-terphenyl in toluene containing 3 g l.$^{-1}$ CCl_4.

From the above results some approximate conclusions can be reached with respect to the time involved in the transfer process. From figure 3 (*a*) it is seen that the addition of CCl_4 at a concentration of 3 g l.$^{-1}$ to a 5 g l.$^{-1}$ solution of terphenyl in toluene quenches the transfer by about 30%. Since, in this solution, the CCl_4 and terphenyl are present at about equal molar concentrations, it follows that the excited toluene molecule transfers its energy to terphenyl in a time which, on the average, is not very much shorter than the average time required for quenching of the excited molecule by CCl_4. It is unlikely that this latter time is appreciably shorter than the mean time which passes before an excited toluene molecule makes a close collision with a molecule of CCl_4. At a concentration of 3 g l.$^{-1}$ this time should be of the order of 10^{-9} second.† It seems therefore that the excited toluene molecule transfers excitation energy to the solute in about 10^{-9} second, a time which is not much shorter than the measured decay times of liquid scintillators which are about 3×10^{-9} second (Post 1950, Swank and Buck 1955).

3.3. *Dependence of Transfer Efficiency on Exciting Wavelength*

The dependence of ϵ on wavelength of the exciting radiation was investigated for terphenyl–toluene solutions at concentrations of 0·1, 1·0, and 5·0 g l.$^{-1}$ in the range between 2200 Å and 2700 Å. It was found that for all three concentrations the intrinsic transfer efficiency is independent of exciting wavelength in the spectral range considered, within a maximum experimental error of about 5%.

In this range the absorption spectra of toluene shows two bands (Friedel and Orchin 1951). For exciting wavelengths shorter than 2300 Å, one therefore expects the second electronic state to be excited. The constancy of ϵ suggests that the transfer always takes place from the first excited state even though the second state may be initially excited. This is consistent with the high probabilities of interval conversion processes in the higher excited states of complex organic molecules and with the rough estimate of the average time for transfer given in the last section.

3.4. *Influence of Dilution of the Transferring Solvent on the Transfer Efficiency*

Among the various ideas which have been proposed in connection with the transfer process, it has been suggested that the solvent molecular excitation energy is propagated through the solvent by a quantum mechanical exchange interaction between neighbouring solvent molecules until an excited solvent molecule finally transfers its energy to a neighbouring solute molecule (Kallmann and Furst 1950). By making certain assumptions concerning interaction energies between neighbouring solvent molecules, Harrison (1951) has arrived at plausible estimates for the time involved in this transfer process, bearing in mind the measured fluorescent decay times in liquid organic solutions, which are of the order of 3×10^{-9} second (Swank and Buck 1955).

In order to investigate this possibility experimentally, the intrinsic transfer efficiency in liquid systems was investigated as a function of the degree of dilution of the transferring solvent by a second solvent transparent to the exciting wavelength, while the solute concentration was kept constant. In this way, if a sufficiently thick layer of solution is used, the number of absorbing solvent molecules which are excited is left unchanged, but, on the other hand, the average distance between the molecules of the absorbing solvent is increased. The function of the second solvent is to separate molecules of the first solvent; in other respects it should be as far as possible inert. According to the above theory of intermolecular propagation through solvent nearest neighbours a decrease of transfer efficiency would be expected with increasing dilution.

---

† In this rough estimate it is assumed that a molecule of toluene makes about 5×10^{11} 'encounters' per second with other molecules. Since molecules in a liquid may make a number of successive collisions with each other before separating, it is necessary to distinguish the total number of collisions, which is of the order of 2×10^{12} per second at room temperature, from the smaller number of 'encounters' or collisions with different molecules (see, for example, Bowen and Wokes 1953).

The results to be described below were obtained for the following systems, in which the absorbing solvent was diluted with cyclohexane: terphenyl in toluene, anthracene in toluene, anthracene in anisole. Cyclohexane was found to be a satisfactory inert and transparent solvent after careful purification. Similar qualitative results were obtained using alcohol as an diluting solvent. Measurements with terphenyl in toluene were carried out up to dilution ratios of 1 : 100 of toluene in cyclohexane. Higher dilution could not be used, because of the prohibitively large corrections arising from direct absorption of the exciting radiation by the solute at 2537 Å. Since the absorption coefficient of anthracene at 2650 Å is about 10% that of terphenyl at 2537 Å, whereas the toluene absorption is unchanged, measurements with anthracene in toluene permit dilutions up to one part in 400 without large corrections. At 2650 Å, the absorption in anisole is about six times greater than in toluene; thus measurements could be made up to dilution ratios of 1 : 1500 in anthracene–anisole systems, without excessively large corrections for direct solute absorption.

The results are shown in figure 4 (*a*), (*b*) and (*c*), using a logarithmic scale for the dilution factor. In all but one of the curves, we see that after a small initial decrease in the transfer efficiency corresponding to relatively low dilution factors, the transfer efficiency remains remarkably constant (within a few per cent) even at the very highest dilutions. Curve 1 of figure 4 (*c*), corresponding to 0·25 g l.$^{-1}$ anthracene in anisole, displays the same general behaviour, but there is an indication of a slight rise after the initial drop.

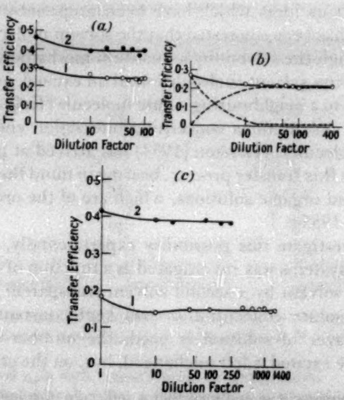


Figure 4. (*a*) Variation of ϵ in solutions of *p*-terphenyl in toluene–cyclohexane with dilution factor. Concentration of *p*-terphenyl: 1, 0·25 g l.$^{-1}$; 2, 0·5 g l.$^{-1}$.
(*b*) Variation of ϵ in solutions of anthracene in toluene–cyclohexane with dilution factor. Concentration of anthracene 0·25 g l.$^{-1}$.
(*c*) Variation of ϵ in solutions of anthracene in anisole–cyclohexane with dilution factor. Concentration of anthracene: 1, 0·25 g l.$^{-1}$; 2, 1 g l.$^{-1}$.

In order to interpret the experimental curves correctly, it is first necessary to consider carefully the influence of dilution on the natural fluorescence of the absorbing solvent, for it is well known that the natural fluorescence of organic solvents increases considerably with dilution, because of the decrease in the extent of self-quenching. Figure 5 shows our experimental results for the natural fluorescence of anisole. Although the present authors (Cohen and Weinreb 1954) have shown that in undiluted systems reabsorption of the solvent fluorescence by the solute is not a process which can account to any significant extent for the observed transfer, it was necessary to examine this point afresh, in the case of diluted systems. This was investigated by comparing the solute fluorescence intensity (isolated from the solvent fluorescence and primary radiation by a suitable filter) obtaining from 'thin' and 'thick' layers of solution, under conditions such that the primary exciting radiation is completely absorbed by the absorbing solvent; here 'thick' and 'thin' implies distances large or small compared with the mean free paths of the natural fluorescent radiation of the absorbing solvent in the solution. The existence of an appreciable radiative tranfser from the solvent to the solute would give rise to an increased solute fluorescence when a thick layer is used, since only in this case is a large fraction of the solvent fluorescence completely absorbed by the solute. It was found that in the case of toluene, the amount of radiative transfer was not greater than 2% of the total transfer, even at high dilution and at the lowest concentrations used in the dilution experiments described above. The natural fluorescence of anisole is, however, appreciably greater than that of toluene, and for a solution of 0·5 g l.$^{-1}$ of anthracene in anisole it is not surprising that the presence of a small amount of radiative transfer was found. This does not exceed 10% of the total transfer, even at high dilutions.

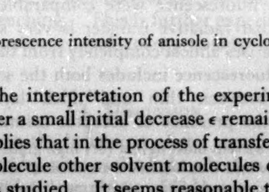


Figure 5. Variation of fluorescence intensity of anisole in cyclohexane with dilution factor.

Returning now to the interpretation of the experimental results shown in figure 4, the fact that after a small initial decrease ϵ remains sensibly constant with increasing dilutions, implies that in the process of transfer from an excited solvent molecule to a solute molecule other solvent molecules can play no part, at least for the diluted solutions studied. It seems reasonable to believe that this is also approximately the situation in the undiluted solutions. The fact that all the graphs in figure 4 show a small initial decrease in ϵ suggests, nevertheless, that in the undiluted systems there exists a small additional possibility of a fast transfer through neighbouring solvent molecules as a consequence of a short range interaction. It is however, possible, in principle, to resolve any of the curves shown in figure 4 into the sum of two curves, for example, such as is shown by the dotted lines in figure 4 (*b*), in which one assumes that at very low dilution the transfer takes place mainly through the solvent molecules according to the proposed theory, while at high dilutions the reverse is true, there being a continuous transition between these two extremes. In the absence of further evidence, however, we believe this to be the less likely of the two possibilities, and that the short range interaction is small in undiluted systems. The small increase of ϵ in curve 4 (*c*) (obtained at the lower concentration of anthracene in anisole) receives a ready explanation on the basis of the appearance of a significant amount of radiative transfer at sufficiently high dilutions, as explained above; the rise occurs at the expected dilutions, taking into account the results shown in figure 5.

It has to be stated here that the results shown in figure 4 seem to be quite contradictory to the conclusions of Moodie and Reid (1952, 1954) which assert that almost no transfer of energy takes place between molecules of different aromatic compounds which are dissolved in an inert solvent.

3.5. *The Influence of Transfer on the Natural Fluorescence of the Solvent*

If one makes the assumption that the process of natural solvent fluorescence competes with the process of transfer of energy to the solute, then the experimentally observed high efficiencies for transfer compared with the low efficiency for natural fluorescence would suggest that the former process is much faster than the latter. One would then expect that the natural fluorescence of the solvent would be partially quenched by the addition of solute molecules, and that the quenching efficiency should be almost equal to the transfer efficiency. This question was investigated experimentally. For this purpose, it was necessary to choose a system in which the absorption of the natural solvent fluorescence by the solute is as small as possible. It is also necessary to choose a solute with a relatively low overall fluorescent efficiency in order to increase the ratio between the solvent fluorescence and the solute fluorescence obtaining after transfer. Anthracene fulfils both these conditions, and in the experiments the two fluorescence intensities were comparable. Experiments were carried out using simple systems of anthracene in toluene, anthracene in anisole† and also diluted systems, in which the natural fluorescence is greater, in accordance with the discussion of the previous section. In each case three different measurements were made. First, the intensity of the fluorescence of the pure solvent was measured, and then the intensity of the fluorescence of the solution was measured using both 'thick' and 'thin' layers of solution (in the sense of the last section). Since the solvent fluorescence and solute fluorescence were comparable in intensity, and since previous results show that radiative transfer is very small, it follows that the thick layer fluorescence arises almost completely from the solvent–solute transfer, whereas the thin layer fluorescence includes both the solvent and solute fluorescence. The difference between 'thin' and 'thick' layer fluorescence gives the solvent fluorescence in the presence of acceptor solute molecules and must be compared with the solvent fluorescent intensity from the pure solvent. As the fluorescence spectrum of the solvent lies in a different spectral region from that of the solute, the existence of the two different fluorescent radiations in the case of thin layers could be further checked by the use of suitable filters. The results are shown in the table.

---

† In this experiment the solvent must have a very high grade of purity. The anisole was distilled three times *in vacuo* and then underwent a twofold distillation in a Todd fractionating column of about 30 theoretical plates. After every distillation the transmissivity and the fluorescence intensity of the distillate was examined under excitation with ultra-violet radiations at 2650 Å. Practically no changes were detected after the second distillation *in vacuo* and no differences at all could be detected between the different fractions of the distillation in the column. No further purification of the toluene used was considered necessary since the measured fluorescence spectrum (obtained by photography) was in complete agreement with that given in the published literature (Pringsheim 1949).

It is seen that in all cases except one no quenching of the solvent fluorescence was observed at all, within the limits of experimental error, as a consequence of the addition of solute molecule. In one case only, 1 g l.$^{-1}$ anthracene in anisole diluted by a factor of 1 : 10 in cyclohexane, a 20% quenching was observed against a transfer efficiency of about 40%.

System	(1)	(2)
0·25 g l.$^{-1}$ anthracene in toluene	25	<6
0·25 g l.$^{-1}$ anthracene in toluene–cyclohexane 1 : 10	20	<3
0·5 g l.$^{-1}$ anthracene in toluene–cyclohexane 1 : 10	30	<5
1 g l.$^{-1}$ anthracene in anisole–cyclohexane 1 : 10	40	20 ± 6

(1) Transfer efficiency (%); (2) degree of quenching of solvent fluorescence by the solute (%).

These results seem to show that the assumption of competing processes stated at the opening of this section is not true for the first three systems, and only partially true for the fourth system (which contains a higher concentration of solute molecules).

Further evidence for this conclusion can be derived from the experimental results of the previous section which are shown in figure 4. For, if we regard the process of natural solvent fluorescence and of energy transfer as competing with each other and with quenching processes in the same molecule, then one should expect that a large change in the fluorescent efficiency of the solvent, following a change in the conditions producing quenching, should be accompanied by a change in transfer efficiency in the same direction. But a comparison of figures 4 and 5 shows that, whereas the fluorescence efficiency of the solvent increases considerably with increasing dilution of the solvent, the transfer efficiency remains constant to a very high degree.

The relatively poor natural fluorescent efficiencies of the solvents used in the present work show that excited solvent molecules are efficiently quenched in the pure solvent, whereas excited solvent molecules which finally transfer very efficiently can exist for times of the order of 10^{-9} second for terphenyl in toluene suffering appreciable quenching. This is further experimental evidence that to a large extent solvent fluorescence and energy transfer take place from different molecular states.

§ 4. DISCUSSION

The similarity between the behaviour of the organic solutions when excited by ionizing radiations and by ultra-violet light which is absorbed by the solvent, and in particular, the similar dependence of light output on the concentration of the solute, indicate that the actual transfer processes are identical in both cases. It has been argued that the transfer mechanism in liquid scintillators is essentially a radiative one, the solute fluorescence being stimulated by the absorption of the solvent fluorescent radiation (Ageno and Querzoli 1952, Birks 1953). The experiments using ultra-violet excitation rule out the possibility of a significant amount of radiative transfer, except in some special dilute systems, such as anthracene in diluted anisole. Even in the latter example, the amount of radiative transfer is not more than about ten per cent of the total transfer[†].

The experiments dealing with the quenching effect of carbon tetrachloride on the transfer process show that a relatively long time must elapse, on the average, before the excitation energy of a solvent molecule is transferred to the solute, at optimum concentrations. This time is of the order of 10^{-9} second, a time which is not short compared to the observed fluorescent decay times (3×10^{-9} second). The experiments with diluted systems show that in these cases the excitation energy is retained by the same solvent molecule during this time. During 10^{-9} second the excited molecule will make numerous collisions and will diffuse until it approaches a solute molecule.[‡] The assumption of such an initial diffusion process may perhaps find support in the well-known fact that higher solute concentrations are needed in plastic scintillators than in liquid solutions in order to obtain equivalent efficiencies. This may be explained by the absence of the pre-transfer diffusion and the consequent increase in the average distance between the solvent and solute molecule over which the interaction responsible for transfer must be operative. It is also reasonable to assume that this general picture of diffusion of the excited solvent molecules for a considerable time, followed by transfer, holds also for the undiluted liquid systems, although in the latter case the results of § 3.4 seem to provide some evidence for an additional relatively weak transfer of excitation energy through neighbouring solvent molecules before transfer to the solute, in accordance with the hypothesis of Harrison.

It is not possible, however, to assume that the solvent–solute transfer takes place, in general, in a close collision occurring directly between the two molecules; this would lead to scintillation decay times greater than those observed (Harrison 1951) and one must assume some kind of long-range interaction between solvent and solute molecules to account for the final transfer.

A theory of a long-range dipole–dipole interaction has been proposed in order to explain the observed energy transfer between dye molecules in solution (Förster 1949). Förster defines a probable interaction distance R_0 which may be calculated empirically from the experimental dependence of transfer efficiency on concentration. From the results shown in figures 1 (*b*) and (*d*) one obtains in this way values of R_0 equal to 38 Å and 29 Å for terphenyl in toluene and anthracene in anisole respectively. The theoretical values of R_0, however, calculated on the basis of the Förster theory from the absorption and fluorescent properties of the molecules involved, turn out to be much smaller—about 15 Å for both terphenyl in toluene and anthracene in anisole. This may not be surprising, since Förster neglects diffusion processes, which is not justified for the solutions used in the present work. The Brownian diffusion distance for terphenyl in toluene, as calculated from the solution viscosity and the 10^{-9} second lifetime for transfer is about 18 Å. Calculation (details to be published later) shows, however, that even when account is taken of the pre-transfer diffusion, the 'empirical' interaction distances still remain larger than the 'theoretical' values. Thus one finds, for example, that any solute molecule which is present within a sphere of radius 38 Å (the empirical interaction distance for terphenyl in toluene) around an excited solvent molecule, has a probability of 50 per cent of finding itself, at some time within the diffusion process, within a sphere of radius 27 Å around the same solvent molecule. This latter distance is still considerably greater than the theoretical range of 15 Å. It must also be noted that the Förster theory assumes that the transfer process competes with the natural fluorescence of the solvent. This indeed is confirmed by experiment for the dye solutions studied by Förster. But the results of § 3.5 show that this is, at most, only partially true, for the systems studied here. It seems, therefore, that although the final transfer process may well be some sort of dipole–dipole interaction, the Förster theory, in its present quantitative form, is not applicable to the solutions used in the work.

[†] The behaviour of these diluted liquid systems bears a phenomenological resemblance to the behaviour of plastic scintillators, for example, anthracene in polystyrene, where about 20–25% of the transfer is radiative (Krenz 1955).

[‡] Preliminary experiments by one of the present authors (A.W.) on the dependence of the transfer efficiency on the viscosity of the solutions provide corroborative evidence that the excited molecule makes a large number of collisions and diffuses appreciably before transfer occurs. The transfer efficiency for low solute concentrations is found to increase with decreasing viscosity; this illustrates nicely the effect of ' encounters '.

The possibility exists of investigating the nature of the transfer mechanism by measuring the degree of polarization of the solute fluorescence after primary excitation of the solvent with polarized ultra-violet light. For a dipole–dipole interaction, some polarization should be detected, but only under very favourable circumstances including high viscosity and high transfer efficiency. Experiments designed to detect such effects are being carried out in this laboratory.

REFERENCES

AGENO, M., CHIOZZOTTO, M., and QUERZOLI, R., 1950, *Phys. Rev.*, **79,** 720.

AGENO, M., and QUERZOLI, R., 1952, *Nuovo Cum.*, **9**, 282.

BIRKS, J. B., 1953, *Scintillation Counters* (London : Pergamon Press).

BOWEN, E. J., 1945, *J. Chem. Phys.*, **13,** 306.

BOWEN, E. J., and WOKES, F., 1953, *Fluorescence of Solutions* (London : Longmans, Green).

COHEN, S. G., and WEINREB, A., 1954, *Phys. Rev.*, **93,** 1117.

FÖRSTER, TH., 1949, *Z. Naturf.*, **4,** 321.

FRIEDEL, R. A., and ORCHIN, M., 1951, *Ultraviolet Spectra of Organic Compounds* (New York : John Wiley).

FUCHSLOCHER, G., and GLASER, G., 1954, *Z. Phys.*, **138,** 363.

FURST, M., and KALLMANN, H., 1954, *Phys. Rev.*, **94,** 503 ; 1955, *J. Chem. Phys.*, **23,** 4, 607.

HARRISON, F. B., 1951, *Thesis*, Princeton University.

KALLMANN, H., and FURST, M., 1950, *Phys. Rev.*, **79,** 857.

KRENZ, H. F., 1955, *Trans. Faraday Soc.*, **51,** 172.

MOODIE, M. M., and REID, C., 1952, *J. Chem. Phys.*, **20,** 1510 ; 1954, *Ibid.*, **22,** 1126.

POST, R. F., 1950, *Phys. Rev.*, **79,** 735.

PRINGSHEIM, P, 1949, *Fluorescence and Phosophorescence* (New York: Intersience).

REYNOLDS, G. T., HARRISON, F. B., and SALVINI, G., 1950, *Phys. Rev.*, **78,** 488.

SWANK, R. K., and BUCK, W. L., 1955, *Rev., Sci. Instrum.*, **26,** 15.

VAVILOV, S. I., 1922, *Phil. Mag.*, **43,** 307.

Single Crystal Bismuth Telluride

By L. AINSWORTH†
Research Laboratories of The General Electric Company, Limited, Wembley, Middx.

MS. received 23rd November 1955

Abstract. Bismuth telluride has been investigated because of its promise as a material for thermo-electric applications. Attempts have been made to grow single crystals of the material by the Stockbarger method, by directional freezing in a horizontal boat, and by pulling from the melt. The first two methods were not found to be satisfactory because of segregation of tellurium towards the last end of the ingot to freeze. This was particularly marked at the slow speeds of withdrawal desirable for good single crystal growth. At fast speeds of withdrawal the segregation was prevented, but at these speeds large single crystals were not produced. Larger crystals, about 1 cm × 2 cm × 3 cm were produced by the pulling method, using a hydrogen atmosphere. These were all p-type of thermo-electric power 200–215 μv deg^{-1} even when considerable departures from the stoichiometric composition Bi_2Te_3 were made in the charge from which the crystals were pulled. The energy gap of this material, measured optically, was 0·15 ev. A p-type single crystal specimen of thermo-electric power 200 μv deg^{-1} had a room temperature electrical conductivity of 500 Ω^{-1} cm^{-1} parallel to the cleavage planes. The room temperature thermal conductivity of the same specimen was 0·0175 w cm^{-1} deg^{-1} parallel to the cleavage planes, and 0·0076 w cm^{-1} deg^{-1} perpendicular to the cleavage planes.

§ 1. INTRODUCTION

In 1910, Haken published results showing the relationship between thermo-electric power and composition in the bismuth–tellurium system. From these results and later work it is seen that the maximum thermo-electric power in the system is obtained in the range round about the composition Bi_2Te_3. A small excess of bismuth in this range leads to p-type material of high thermo-electric power, while a larger bismuth excess yields n-type material. A tellurium excess produces n-type material of larger thermo-electric power than the n-type bismuth rich compositions. Some interest has been shown recently in the semiconductor bismuth telluride Bi_2Te_3, as a possible material for thermo-electric applications. This material has a high ratio of electrical to thermal conductivity, a factor of importance in these applications (Goldsmid and Douglas 1954), and work has been done in these laboratories on this material. Preliminary work showed that thermo-electric powers up to 240 μv deg^{-1} could be obtained, using very pure starting materials, and that the material usually produced on melting down a charge weighed to give the stoichiometric composition Bi_2Te_3 was p-type n-type material of similarly high thermo-electric power was produced with difficulty, by adding considerable excesses of tellurium (up to 5%) to a charge of the stoichiometric composition and passing a molten zone down the charge at a rate of about ½ in. hr.$^{-1}$ Chemical and x-ray analyses showed no difference in composition between p-type material of maximum thermo-electric power, intrinsic material, and n-type material of maximum thermo-electric power, indicating that the change from maximum p-type to maximum n-type material occurred within a composition range of 0·2%, the limit of sensitivity of the analyses. This is in agreement with the findings of Vasenin (1955). Furthermore it can be predicted that the change from maximum p-type to maximum n-type material should occur within this composition range. It is evident, therefore, that only a small fraction of the considerable excess of tellurium necessary to produce maximum n-type material is, in fact, going into the main crystal structure, the unincorporated material remaining as a separate phase, possibly at the grain boundaries. To overcome the possibility of grain boundary phases in the polycrystalline ingots, it was decided to try to grow single crystals of bismuth telluride.

---

† Now at Department of Glass Technology, University of Sheffield.

§ 2. THE STOCKBARGER METHOD

The first method used was an adaptation of the Stockbarger (1936) method, the apparatus being as shown in figure 1. The two furnace windings were independently controlled, to ±1°c, by means of platinum resistance type controllers. The upper and lower furnaces were separated by means of a platinum baffle and hardened steel liners were used to render the temperatures in each furnace fairly uniform. The gap between the upper and lower furnaces was 1 in. and the temperature difference between the two was usually in the range 100–125°c.

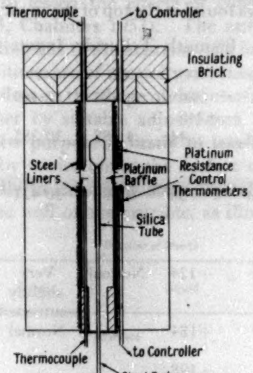


Figure 1. The Stockbarger-type furnace.

The raw materials used were zone refined to a very high degree of purity, such that no impurities were detected spectrographically in the bismuth, and only a trace of bismuth was present as impurity in the tellurium. They were carefully weighed and sealed inside pointed bottom silica containers, evacuated to a pressure of less than 10^{-5} mm Hg. The containers were supported as indicated, by means of a silica tube on the end of a metal rod. The upper end of the rod was in contact with the point of the silica crucible, and the lower end passed through a water-cooled jacket. This was to ensure heat flow in the right direction as the charge was lowered from the hot to the cold part of the furnace.

The speed of withdrawal was varied between 2 in. hr^{-1} and 1/10 in. hr^{-1}. It was found that at the slow rates of withdrawal desirable for good single crystal growth, segregation of tellurium to the top of the ingot took place. This free tellurium was detected by x-rays, and in the case of the slowest speeds of withdrawal was present in sufficiently large amounts to cause detectable differences in the lattice parameters of the bulk, bismuth-rich, bismuth telluride. In particular, appreciable contractions in the c_0 lattice parameter were detected, i.e. there were marked decreases in the spacings of the (0001) cleavage planes and slightly larger changes in the axial ratio c/a of the hexagonal structure cell (see table). Electrical measurements on these tellurium deficient ingots showed them to be n-type of low thermo-electric power (20–50 μv deg^{-1}), in qualitative agreement with the curves published by Haken and Vasenin, i.e. a considerable excess of bismuth causes the material to become n-type. This excess bismuth is taken up in the main crystalline phase, as shown by the x-ray examination. No similar tellurium excesses could be detected in the main crystalline phase of any of the ingots by x-ray examination, even when excess tellurium was added to the starting charge in attempting to produce n-type material of high thermo-electric power. The excess tellurium was always found at the top of the ingot, as free tellurium. When

Bismuth Telluride Ingots

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
†Normal Bi_2Te_3	See below	—	+200 to 240	4·384	30·469	6·95	
B.T.A1	Vertical furnace	1/10	−44	4·388	30·385	6·925	Ingot consisted of 3 good quality single crystals.
B.T.A6	Horizontal furnace	1/10	−58	4·385	30·421	6·94	Two large single crystals and one or two smaller ones.
B.T.A9	Vertical furnace	1/6	+124	Normal	Very slightly contracted	Almost normal	Crystal size not large.
B.T.A12	,,	1/2	+184	,,	Normal	Normal	Crystal size fairly small.
B.T.A17	,,	1	+198	,,	,,	,,	Crystal size small
B.T.A20	,,	2	+214	,,	,,	,,	Crystal size very small.
B.T.A19	Pulling method	0·2‡	+218	,,	,,	,,	Single crystal, size 1 cm × 2 cm × 3 cm.

† This is the material obtained by simply melting down pure bismuth and tellurium to the stoichiometric composition inside an evacuated sealed container. The lattice parameters quoted agree with those given in the A.S.T.M. index.

‡ (mm min^{-1}).

(1) Source of specimen; (2) method of preparation; (3) speed of withdrawal (in. $hr.^{-1}$); (4) thermo-electric power (μv deg^{-1}); (5)–(7) lattice parameters: (5) *a* direction, (6) *c* direction, (7) c/a; (8) remarks.

the speed of withdrawal was increased to $\frac{1}{2}$ in. hr^{-1} or more there was very little segregation of tellurium to the top surface, and stoichiometric Bi_2Te_3 charges yielded p-type material of nearly the maximum thermo-electric power. At these fast rates of withdrawal, however, the ingots were by no means single crystal. A summary of some of the results is given in the table.

This method is therefore not a good one for obtaining single crystals of bismuth telluride. At the slow rates of withdrawal segregation of tellurium to the top of the ingot invariably occurs. This is particularly noticeable when excess tellurium is added to try to produce high thermoelectric power n-type material, but also occurs when charges of the stoichiometric composition Bi_2Te_3 are used. At the faster rates of withdrawal the crystal size is not good.

§ 3. HORIZONTAL DIRECTIONAL FREEZING

Attempts were also made with the furnace in a horizontal position, the charge being withdrawn horizontally from the hot end to the cold end. This was done to allow a large free top surface, to take up any strains due to expansion differences between charge and boat. The important conditions to be fulfilled in attempting to grow single crystals by this method, as for the previous method, are (i) pure starting materials, (ii) steep temperature gradient across the liquid–solid interface, (iii) slow rate of traverse, (iv) a liquid–solid interface which is convex as viewed from the liquid.

The first condition is to lessen the effects of constitutional supercooling (Rutter and Chalmers 1953, Chalmers 1953). The next two conditions are to lessen the effect of supercooling due to the existence of a finite rate of freezing, and hence the fact that the conditions are of necessity not those of equilibrium. The third condition has also an effect on the shape of the interface. These three conditions can be, and were, met by suitable adjustments. The fourth condition presents more difficulty. Its purpose is firstly to lessen the possibility of the formation of stray crystals by nucleation at the surface of the boat ahead of the advancing solid and secondly to cause any stray crystals which might form to grow out immediately to the wall of the crucible, as illustrated in figure 2. In figure 2 (*a*) the temperature at the wall of the boat near the solid– liquid interface is higher than that in the centre of the charge. There will therefore be less chance of a stray crystal nucleating at the wall ahead of the advancing solid than there will be in case 2 (*b*), where the temperature at the wall is lower than that in the centre of the

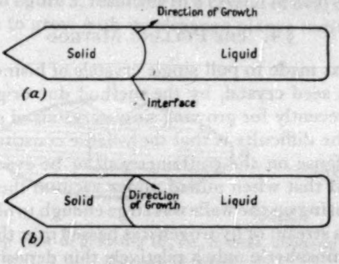


Figure 2. Illustrating the importance of the shape of the interface on the growth of single crystals.

charge. Furthermore, in case 2(*a*) if a stray crystal does form at the liquid–solid interface it will usually grow out to the side wall immediately, whereas in 2(*b*) a stray crystal which nucleates will grow wider as growth continues, as indicated by the direction of the arrow.

The factor which governs the shape of the interface is the ratio of the heat lost by conduction from the interface along the solid charge, to that lost laterally through the sides of the boat. The rate of growth of the crystal depends on the rate of loss of the latent heat of solidification from the interface, and for the right shape of interface most of this heat should be lost by conduction through the solid charge to the cold end of the ingot and thence by radiation and conduction through the supporting rod. To fulfil this condition ideally the charge should have a very much higher thermal conductivity than the boat material, and this is invariably the case with metals, which are easily grown by this method. Semiconductors generally are more difficult to grow in single crystal form by this method, because of their lower conductivities. Bismuth telluride in particular is difficult, because low thermal conductivity is one of its advantages for thermoelectric applications. Figure 3 (Plate) illustrates this point very well. This is a photograph of an ingot of Bi_2Te_3 grown by this method at a speed of $\frac{1}{8}$ in. hr^{-1}. The ripple marks, which may be caused by tapping the supports lightly during freezing, indicate the shape of the interface and it can be seen that in this case the interface is of the wrong shape, being concave as viewed from the liquid. The resulting crystals growing in along parallel lines from the side of the boat may be seen towards the last end of the ingot to freeze. The ingot has been etched to reveal the individual crystals more clearly. This has rendered the ripple marks fainter than on the untreated ingot; they can, however, be seen sufficiently well to indicate the shape of the interface.

Again there was segregation of tellurium towards the last end to freeze when slow speeds of withdrawal were used to promote larger single crystal growth. A small tellurium end cap can in fact be seen on the ingot shown in figure 3. In both these methods therefore slow speeds of withdrawal led to segregation of tellurium, while faster speeds were not conducive to [illegible] single crystal growth.

§ 4. THE PULLING METHOD

Attempts were next made to pull single crystals of bismuth telluride from the melt, on the end of a seed crystal, by the method due originally to Czochralski (1918) and modified recently for growing single crystals of germanium (Teal and Little 1950). Here the difficulty is that the volatile constituent, tellurium, tends to come off and condense on the container wall or be evacuated in the vacuum pumps. It was found that when pulled under vacuum the amount of tellurium coming off and condensing on the walls was large enough to obscure the view after a short while, but when a stream of hydrogen was passed over the charge at a pressure very slightly above atmospheric only a relatively thin deposit of tellurium formed during a run. The system finally used was as indicated in figure 4. Heat was supplied by means of a molybdenum strip furnace, the energy input being controlled by adjustments to two continuously variable voltage supplies in series, one a coarse control and the other, adjustable to 1/20 volt, a fine control. Both supplies were obtained from a stabilized constant voltage source. Hydrogen was passed continuously through the tube containing the charge. The seed crystal was rotated continuously during the run and withdrawn by means of an electric motor operating a screw drive through a variable speed gear box.

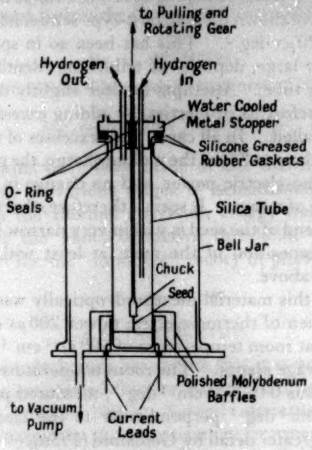

to Pulling and Rotating Gear
Hydrogen Out
Hydrogen In
Water Cooled Metal Stopper
Silicone Greased Rubber Gaskets
O-Ring Seals
Silica Tube
Bell Jar
Chuck
Seed
Polished Molybdenum Baffles
to Vacuum Pump
Current Leads

Figure 4. The pulling apparatus.

It is necessary to cut the seed crystals with the cleavage planes running lengthwise down the seed. This is the only way in which the seed is mechanically strong enough to support a crystal growing on the end. It cleaves readily parallel to the basal planes of the hexagonal structure cell. This being so, it might be expected that the crystal growing on the seed would tend to grow at different speeds in the different crystallographic directions. This was found to be so. The crystal tended to grow outwards more quickly in the direction parallel to the cleavage planes, direction AA in figure 5, resulting in a crystal of oval cross section. Often, also, the crystal tends to grow with its cleavage planes radial, in directions other

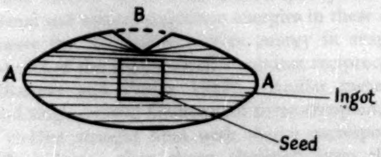

B
A
A
Ingot
Seed

Figure 5. Schematic top view of growing Bi_2Te_3 crystal, showing the relative orientations of seed and ingot.

than AA, and when this happens a V forms down the middle of one side of the crystal, the bisector of the V being perpendicular to the direction of the cleavage planes in the seed crystal, as indicated at B in figure 5. This is a characteristic feature of crystals grown by this method and figure 6 (Plate) shows photographs of two typical ingots, showing the V down one side. The central portions of such crystals, are however, usually of good single crystal material. The departure from single crystal formation is easily detected by simple visual examination.

A number of single crystals of Bi_2Te_3 have been made by this method. In all cases measurements have shown them to be p-type semiconductors of high thermo-electric power, 200–220 μv deg^{-1}. This has been so in spite of very noticeable, though not extensively large, deposits of tellurium, identified by x-rays, on the walls of the containing tube. Attempts to alter slightly the composition of the pulled crystal, and therefore its properties, by adding excess bismuth or tellurium to the starting charge failed. In all cases, with excesses of tellurium up to 5% or with excesses of bismuth up to 2% in the molten charge, the pulled ingot was p-type of the same high thermo-electric power, and no change in the lattice parameters was detected by the use of x-rays. It seems, therefore, that the composition of the crystal growing on the end of the seed is within very narrow limits always the same irrespective of the composition of the melt, at least within the range of melt compositions indicated above.

The energy gap of this material measured optically was found to be 0·15 ev. A single crystal specimen of thermo-electric power 200 μv deg^{-1} (p-type) had an electrical conductivity at room temperature of 500 Ω^{-1} cm^{-1}, measured in a direction parallel to the cleavage planes. The room temperature thermal conductivity of the same specimen was 0·0175 w cm^{-1} deg^{-1}, measured parallel to the cleavage planes, and 0·0076 w cm^{-1} deg^{-1} perpendicular to the cleavage planes. These results are reported in greater detail by Goldsmid (1956).

REFERENCES

CHALMERS, B., 1953, *Canad. J. Phys.*, **31**, 132.
CZOCHRALSKI, J., 1918, *Z. Phys. Chem.*, **92**, 219.
GOLDSMID, H. J., 1956, *Proc. Phys. Soc.* B, **69**, 203.
GOLDSMID, H. J., and DOUGLAS, R. W., 1954, *Brit. J. Appl. Phys.*, **5**, 386.
HAKEN, W., 1910, *Ann. Phys., Lpz.*, **32**, 319.
RUTTER, J. W., and CHALMERS, B., 1953, *Canad. J. Phys.*, **15**, 1531.
STOCKBARGER, D. C., 1936, *Rev. Sci. Instrum.*, **7**, 133.
TEAL, G. K., and LITTLE, J. B., 1950, *Phys. Rev.*, **78**, 647.
VASENIN, F. I., 1955, *Zh. Tekh. Fiz.*, **25**, 397.

Some Observations on the Theory of Photoconductivity in the Lead Sulphide Group of Compounds

By C. WOOD†

The Plessey Co. Ltd., Caswell Research Laboratories, Towcester, Northants.

MS. received 16th *January* 1956

Abstract. Measurements have been made of the thermal activation energy of PbSe evaporated films at various stages of sensitization treatment. For the highly sensitized film the thermal activation energy was found to be 0·29 ev, which corresponds to the photoconductive threshold. This result has been interpreted as support for the ' numbers ' theory of photoconductivity.

§ 1. INTRODUCTION

PREVIOUS experimental work has shown that it is necessary to expose PbS, PbSe and PbTe evaporated films to oxygen in order to obtain appreciable photosensitivity (Smith 1951). Several attempts have been made to explain the role of oxygen in the sensitization treatment, and the theories of photoconductivity in these compounds fall into two main categories. The first, which may be called the barrier theory, supposes that Shottky-type potential barriers are present in the layer and that radiation lowers the height of these barriers and thus increases the mobility of the current carriers. According to this theory the role of oxygen is to produce these potential barriers (Sosnowski *et al.* 1946, James 1949, Rittner 1950, Chasmar 1948, Chasmar and Putley 1951).

Under the second class come the normal or 'numbers' theories of photoconductivity in which radiation increases the number of current carriers and has negligible effect on the mobility. The strongest evidence for this theory is that Hall effect measurements on PbTe (Levy 1953) have shown that a large increase in numbers of carriers occurs on illumination, whereas the change in mobility is small. However, the role of oxygen in this case has appeared to be more obscure. This was mainly because of a discrepancy between the measured values of the thermal and optical activation energies in these compounds. Early attempts to measure the thermal activation energy in evaporated films failed because the usual plot of the log (resistance) against reciprocal temperature gave curved lines (Chasmar and Putley 1951). Similar measurements on large polycrystalline and single-crystal filaments of these compounds at high temperatures (>500°K) yielded straight lines with slopes corresponding to activation energies very much larger than those obtained from the photoconductive threshold (Putley 1952).

Scanlon (1953, 1954) was able to make measurements at much lower temperatures by using very pure single crystals, and he obtained thermal activation energies which correspond closely to the optical values. He suggested that both atomic and electronic changes are taking place at high temperatures and therefore the values obtained by Putley are invalid.

---

† Now at Philco Corporation Research Division, Philadelphia, Pennsylvania.

§ 2. Experimental Procedure and Results

It is the purpose of this paper to show that the optical and thermal activation energies in sensitized evaporated films are in fact in agreement down to temperatures well below room temperature (~ 180°K). New measurements have been made of the variation of resistance with temperature of several PbSe evaporated photocells. In order that only electronic changes should take place, these measurements have been carried out in the temperature range from room temperature down to the boiling point of liquid nitrogen. The photocell was placed in a dark evacuated enclosure inside a Dewar vessel containing liquid nitrogen. The temperature of the photocell was measured by a chromel–alumel thermocouple attached to one of the electrodes.

Figure 1(*a*) shows the results obtained on a highly sensitized PbSe photocell. This cell was sensitized by baking in air, and possessed a limiting sensitivity at room temperature of about 10^{-10} watt incident on a sensitive area of 0·15 cm^2 for 1 c/s bandwidth and 800 c/s chopped radiation of 3 μ wavelength.

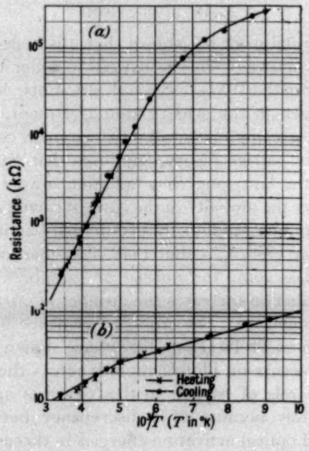


Figure 1. (*a*) Fully sensitized photocell. (*b*) Photocell exposed to air.

It is seen that a straight line is obtained down to quite low temperatures. From the slope of this curve an activation energy of 0·33 ev is obtained. This value is in good agreement with the photoconductive threshold for this cell (~ 3·5 μ), i.e. the value at which the sensitivity has fallen to one-half of the peak value. Other highly sensitized cells have yielded similar activation energies. However, in plotting log (resistance) against $10^3/T$°K to determine the activation energy it is assumed that the mobility is varying according to a $T^{-3/2}$ law. It has been shown that, for the PbS type compounds, the mobility μ in single crystals (Putley 1952) obeys a law of the form $\mu \propto T^{-5/2}$. If this law is obeyed in evaporated films of PbSe, the ordinates of figure 1(*a*) should be divided by the absolute temperature T. From this curve, an activation energy of 0·29 ev is obtained.

We have attempted to sensitize PbSe photocells without baking, by exposing the film to various pressures of air at room temperature, but the photosensitivity at room temperature has been negligible. However, exposure of a sensitized cell to air at room temperature destroys the photosensitivty and for this reason the evaporated film in our PbSe photocells is situated in an evacuated Pyrex enclosure ($\sim 10^{-5}$ mm Hg). After the measurements described above had been made the photocell was broken open and exposed to air at room temperature for 24 hours. This treatment destroyed the photosensitivity and caused the dark resistance to fall by a factor of about 25. The variation of resistance was again measured and the results are plotted in figure 1(*b*). In this case the curve consists of two straight lines with slopes of 0·05 ev and 0·025 ev. The results of measurements on a PbSe film evaporated in a vacuum of less than 10^{-5} mm Hg with no subsequent sensitization treatment, and therefore possessing very little photosensitivity, is shown in figure 2(*a*). The slope of the curve corresponds to 0·015 ev. Curves are also shown of cells which had been over-sensitized by baking in air (figures 3(*a*) and (*b*)) and of a cell which had been sensitized by baking treatment but not sufficiently to reach maximum sensitivity (figure 2(*b*)). In the latter case the curve is non-linear and therefore the activation energy could not be calculated.

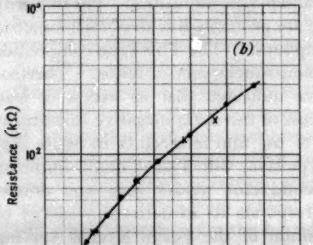


Figure 2. (*a*) Non-sensitized photocell. (*b*) Partially sensitized photocell.

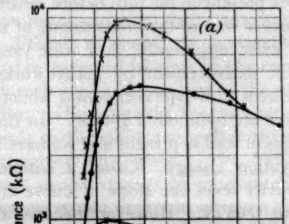


Figure 3. (*a*) Over-sensitized photocell. (*b*) Greatly over-sensitized photocell.

Only non-linear curves were obtained by Chasmar and Putley (1951) and by Simpson (1947) from their results on evaporated films of PbTe and PbSe. Our results suggest that this was due to their films being insufficiently sensitized.

Simpson postulated that the resistance–temperature characteristics could be explained by a variation of the resistance of the intercrystalline contacts of the film, due to the differential coefficient of expansion of the film and the glass substrate. Since our PbSe films were evaporated on to Pyrex glass substrates it was necessary to assess the importance of the differential coefficient in the above results.

A PbSe film was evaporated in a bell-jar apparatus on to a fluorspar (CaF_2) substrate. The linear coefficients of expansion of CaF_2 and PbSe are both approximately 20×10^{-6}. This film was highly sensitized by baking in air. A plot of resistance–temperature measurements in the range from room temperature up to 100°c gave a straight line curve of the same slope as in figure 1(*a*). Therefore, it was concluded that the activation energies obtained with Pyrex substrates are an intrinsic property of the PbSe compound, and not due to differential expansion coefficients.

Figures 3(*a*) and (*b*) show that a permanent change in resistance had occurred after one temperature cycle. It is possible that this change is due to cracks in the film caused by the differential coefficients of expansion. This has not however, affected the slope of the curves in figure 3(*a*) at the higher temperatures.

§ 3. DISCUSSION OF RESULTS

The results described here appear to show that one of the main roles of oxygen is to ' purify ' the compound by balancing n- and p-type impurities. The impurity levels present in the film before sensitization are probably due to a stoichiometric excess of one of the constituents of the compound. The PbS-type compounds can exist, apparently, in a wide range of non-stoichiometric ratios. Thermo-e.m.f. measurement by many workers have indicated that an excess of lead produces an n-type compound whilst an excess of the electronegative constituent makes the compound p-type. In the case of the film referred to in figure 2(*a*) excess of lead is present and we have an n-type layer with a small donor-impurity activation energy. Using a simple semiconductor model with one donor-impurity level the slope of curve 2(*a*) corresponds to an activation energy of about 0·03 ev. However, because the position of the Fermi level is varying over this temperature range it is necessary to assume a fairly large concentration of impurities of the order of 10^{18} cm^{-3} in order to obtain an approximately straight line. It is of interest to calculate the value of the impurity activation energy assuming a hydrogen-like model. This energy is given by

$$E_{\mathrm{I}} = \frac{2\pi m^* e^4}{K^2 h^2} = \frac{13{\cdot}5 m^*}{K^2}.$$

Avery (1953) has obtained a value of 4·6 for the refractive index of PbSe at a wavelength of 3 microns; hence the dielectric constant K is about 21. Assuming the effective mass m^* is equal to the free electron mass, $E_{\mathrm{I}} = 0{\cdot}03$ ev. Such good agreement is probably fortuitous for the following reasons: (*a*) it is unlikely that the effective mass is equal to the free electron mass, (*b*) the value for the dielectric constant should lie between the values for high frequency and static fields, and (*c*) 10^{18} impurities cm^{-3} is probably too high a concentration to give the donor depth corresponding to infinite dilution, for which the hydrogen atom approximation is valid.

On exposure to sufficient oxygen, an evaporated film can be changed from n-type to p-type and becomes neutral near peak photosensitivity. Thus, during the initial stages of the sensitization treatment acceptor impurities are introduced so that both types of impurities are present in fairly large, although unequal, numbers and the resistance increases. This state of the layer can give rise to a curved log (resistance) against $1/T$ plot (Shockley 1950). This is the state of the film referred to in figure 2(*b*).

When fully sensitized, the film increases its resistance still further, and becomes intrinsic down to quite low temperatures as in figure 1(*a*), by balance of n- and p-type impurities. It is unlikely that the acceptor and donor impurity activation energies are equal. However, unless the difference is large, intrinsic conduction will still be obtained in the temperature range shown in figure 1(*a*). For example, with an acceptor level 0·06 ev above the full band, and a donor level 0·03 ev below the conduction band, the semiconductor will be intrinsic down to 170°K.

Over-oxidation introduces an excess of p-type impurities with a resultant decrease in resistance, as in figures 3(*a*) and (*b*). It is possible that the actual oxidation mechanism is the replacement of F′-centres by O^{2-} ions as suggested by Ehrenberg and Hirsch (1951).

§ 4. CONCLUSIONS

The barrier theory has been invoked in order to explain the variation of the room-temperature resistance and the thermoelectric behaviour during sensitization (Sosnowski *et al.* 1947). There is little doubt that potential barriers are present in evaporated films of the PbS-type compounds but their contribution to the photoconductive mechanism is more difficult to assess. It is possible that the photoconductive behaviour is influenced by potential barriers and by increases in numbers of carriers, the degree to which each applies depending on the method of preparation and sensitization of the film. However, the thermal activation energy of our evaporated films corresponds to the photoconductive threshold, and to the optical activation energy determined by Gibson (1950, 1952) and Avery (1953, 1954) from optical absorption measurements on evaporated films and single crystals. On the barrier theory one would expect to obtain a thermal activation energy of nearly twice the intrinsic energy gap. Rittner (1950) has suggested a model in which there is a quasi-intrinsic region between the barriers. Our results suggest that if barriers are present in our sensitive layers, then these intrinsic regions must make up the bulk of the layer. Hence the 'numbers theory' appears to be the more likely explanation. Consideration of a simple definition of d.c. photosensitivity S in the 'numbers theory', e.g. $S=(i_{\mathrm{I}}-i_{\mathrm{d}})/i_{\mathrm{d}}$ where i_{I} and i_{d} are the conductances under illumination and in the dark, respectively, will show that a small increase in dark resistance will greatly increase the photosensitivity.

It is immediately obvious why the photoconductive response is many orders of magnitude less in single crystals of the PbS-type compounds than in evaporated films, because of the difficulty in producing single crystals that are intrinsic below room temperature and which are, from optical absorption considerations, of the same order of thickness as evaporated films, i.e. about 1 micron.

One important set of experimental results has been neglected so far. This is the variation of the shape of the spectral response curves with degree of oxidation. Bode and Levinstein (1954) have accounted for this variation in PbTe by assuming that adsorbed oxygen removes electrons from lower levels of the conduction band. It appears likely, however, that oxygen also produces trapping levels in the compound, since they observed a large difference in response time between the long- and short-wavelength peaks in the spectral response curve. An alternative explanation is obtained by postulating an overlapping band structure (Gibson 1950, 1952). Further experiments are planned in our laboratory in order to give more information about the true mechanism.

ACKNOWLEDGMENTS

Thanks are due to Mr. P. Cotterrell for the production of the various photocells, to Mr. P. Herman for the production of the bell-jar film, to Mr. I. M. Ross, of the Services Electronic Research Laboratory for helpful suggestions, and to the Plessey Company for permission to publish this paper.

REFERENCES

AVERY, D. G., 1953, *Proc. Phys. Soc.* B, **66**, 133; 1954, *Ibid.*, **67**, 2.
BODE, D. E., and LEVINSTEIN, H., 1954, *Phys. Rev.*, **96**, 259.
CHASMAR, R. P., 1948, *Nature, Lond.*, **161**, 281.
CHASMAR, R. P., and PUTLEY, E. H., 1951, *Semiconducting Materials*, Ed. H. K. Henisch (London : Butterworths Scientific Publications).
EHRENBERG, W., and HIRSCH, P., 1951, *Proc. Phys. Soc.* B, **64**, 700.
GIBSON, A. F., 1950, *Proc. Phys. Soc.* B, **63**, 756; 1952, *Ibid.*, **65**, 378.
JAMES, H. M., 1949, *Science*, **110**, 254.
LEVY, J. L., 1953, *Phys. Rev.*, **92**, 215.
PUTLEY, E. H., 1952, *Proc. Phys. Soc.* B, **65**, 388, 736, 992.
RITTNER, E. S., 1950, *Science*, **111**, 685.
SCANLON, W. W., 1953, *Phys. Rev.*, **92**, 1573 ; 1954, *Physica*, **20**, 1090.
SHOCKLEY, W., 1950, *Electrons and Holes in Semiconductors* (Princeton : van Nostrand).
SIMPSON, O., 1947, *Nature, Lond.*, **160**, 791.
SMITH, R. A., 1951, *Semiconducting Materials*, Ed. H. K. Henisch (London : Butterworths Scientific Publications).
SOSNOWSKI, L., STARKIEWICZ, J., and SIMPSON, O., 1946, *Nature, Lond.*, **158**, 28 ; 1947, *Ibid.*, **159**, 818.

The Density of Electronic States in Body-Centred Cubic Crystals

By M. A. E. NUTKINS†
Imperial College, University of London

Communicated by H. Jones; *MS. received* 9*th January* 1956

Abstract. The first six Brillouin zones for a metal with a body-centred cubic structure are described in detail. The density of states for each of these zones is found by assuming the electrons to be free. The discontinuities in the energy at points of high symmetry in the Brillouin zone structure are calculated by the approximation of nearly free electrons, and numerical values are obtained for sodium by comparison with the results of Howarth and Jones. The separate $N(E)$ curves for each of the zones are modified in accordance with these values of the energy and the modified total curve is obtained.

§ 1. INTRODUCTION

THE density of states of the conduction electrons in a metal, usually denoted by $N(E)$, is of interest for the interpretation of many of the metal's physical properties. For example, the soft x-ray emission is determined largely by the density of states in the region below the Fermi limit, and the soft x-ray absorption by the density of states in the unoccupied region above this limit. Other properties such as the electronic specific heat and the magnetic susceptibility depend upon the density of states in the immediate neighbourhood of the Fermi limit. For real metals the function $N(E)$ is difficult to calculate; only when the conduction electrons are assumed to be completely free can it be obtained easily. It is then proportional to $\sqrt{E}$.

The present calculation is made for a metal having a body-centred cubic structure. As a first step in this calculation, the free electron approximation has been used to obtain the density of states for each of the first six separate Brillouin zones of the lattice. These curves are plotted in figure 1, and it can be seen that addition of the separate curves reproduces the free electron parabola. In order to obtain a more realistic curve, which takes into account the periodic field of the lattice, the energies at points of high symmetry in the Brillouin zone structure have been calculated formally by using the approximation of nearly free electrons. Since it is well known that this approximation cannot be used to obtain reliable values of the energy gaps by a direct calculation of the Fourier coefficients, these coefficients have been obtained, for the special case of sodium, by a comparison with the energy gaps calculated by the cellular method (Howarth and Jones 1952). The separate $N(E)$ curves, determined for free electrons, have been modified by displacing horizontally, to the energy values appropriate to the real metal, those points on the curves which correspond to points of high symmetry in the Brillouin zones. The whole of each separate $N(E)$ curve has then been obtained by proportional displacements of each point. Finally the ordinates have been scaled so that the area of each curve remains unaltered. The total density of states is given by the addition of the separate curves. This method, while admittedly approximate, should give a good idea of the general form of the density of states, and has the advantage that it can be applied to large energies where detailed calculations would become impracticable.

---

† Now at Bedford College, University of London.

§ 2. GEOMETRY OF THE FIRST SIX BRILLOUIN ZONES

The Brillouin zones for any cubic lattice, of lattice constant a, are formed by the planes whose equations are

$$\mathbf{n}.\mathbf{k}=\pi n^2/a, \qquad \ldots\ldots(1)$$

where $\mathbf{k}$ is the wave vector defined so that $\hbar\mathbf{k}$ is the momentum of the electron, and $\mathbf{n}$ is a vector with integer components (n_1, n_2, n_3). In the particular case of the body-centred cubic lattice the energy does not change discontinuously across all of these planes ; only those for which the structure factor $1+\exp \pi i(n_1+n_2+n_3)$, does not vanish need to be considered.

In order to simplify the description of points in the Brillouin zones it is convenient to change the scale of $\mathbf{k}$ and to let

$$\mathbf{k}'=(a/\pi)\mathbf{k}. \qquad \ldots\ldots(2)$$

In the new units equation (1) has the form

$$\mathbf{n}.\mathbf{k}'=n^2. \qquad \ldots\ldots(3)$$

The first six zones have been determined by equation (3), together with the rule that a line joining the origin to any point in the nth zone crosses exactly $n-1$ planes of finite energy discontinuity.

The following is a description of the zones, in which it should be noted that the coordinates are referred to axes whose scale is determined by equation (2) :

First Zone. The first zone consists of the dodecahedron formed by the set of planes $\{1, 1, 0\}$.

Second Zone. The outer boundaries of the second zone are formed by extending the $\{1, 1, 0\}$ planes. The zone consists of twelve pyramids whose bases are the faces of the dodecahedron. The coordinates in k'-space of the vertices of a typical pyramid are (1, 1, 1), (−1, 1, 1), (2, 0, 0), (0, 2, 0), (0, 2, 2).

Third Zone. The outer boundaries of the third zone are formed by the further extensions of the $\{1, 1, 0\}$ planes together with the set of planes $\{2, 0, 0\}$. The zone consists of twenty-four tetrahedra ; a typical tetrahedron having vertices at the points (2, 0, 2), (0, 2, 2), (0, 0, 2), (1, 1, 1).

Fourth Zone. The outer boundaries of the fourth zone are formed by the extended $\{1, 1, 0\}$ and $\{2, 0, 0\}$ planes together with the $\{2, 1, 1\}$ planes. The elements of the zone may be divided into two sets which are such that the elements of each set are similar and similarly situated with respect to the origin, and so may be taken as identical for the purposes of the present calculation. The first set consists of twenty-four tetrahedra, a typical tetrahedron having vertices at the points (0, 0, 2), (2, 0, 2), (0, 2, 2), $(\frac{1}{2}, \frac{1}{2}, 2\frac{1}{2})$. The second set consists of eight figures, each made up of two tetrahedra with the same base, a typical example having a base determined by the points (0, 2, 2), (2, 0, 2), (2, 2, 0), and vertices by the points (1, 1, 1) and $(\frac{3}{2}, \frac{3}{2}, \frac{3}{2})$.

Fifth Zone. The outer boundaries of the fifth and sixth zones are formed by the extensions of all the planes previously considered. The fifth zone has three sets of elements : (*a*) twenty-four pyramids with quadrilateral bases, a typical pyramid having as base the quadrilateral specified by the points $(0, 0, 2)$, $(\frac{1}{2}, \frac{1}{2}, 2\frac{1}{2})$, $(-\frac{1}{2}, \frac{1}{2}, 2\frac{1}{2})$, $(0, \frac{2}{3}, 2\frac{2}{3})$ and the vertex by the point $(0, 2, 2)$, (*b*) twenty-four tetrahedra of the type whose vertices are given by $(\frac{1}{2}, \frac{1}{2}, 2\frac{1}{2})$, $(2, 0, 2)$, $(0, 2, 2)$, $(1, 1, 3)$ and (*c*) twenty-four tetrahedra of the type with vertices $(2, 0, 2)$ $(0, 2, 2)$, $(\frac{4}{3}, \frac{4}{3}, 2)$, $(\frac{3}{2}, \frac{3}{2}, \frac{3}{2})$.

Sixth Zone. The sixth zone has four sets of elements : (*a*) forty-eight tetrahedra of the type with vertices $(\frac{1}{2}, \frac{1}{2}, 2\frac{1}{2})$, $(0, \frac{3}{2}, 2\frac{2}{3})$, $(1, 1, 3)$, $(0, 2, 2)$, (*b*) twenty-four tetrahedra of the type with vertices $(1, 1, 3)$, $(0, 2, 2)$, $(2, 0, 2)$, $(\frac{4}{3}, \frac{4}{3}, 2)$, (*c*) twenty-four pyramids of the type with base $(\frac{4}{3}, \frac{4}{3}, 2)$, $(\frac{4}{3}, 2, \frac{4}{3})$, $(\frac{3}{2}, \frac{3}{2}, \frac{3}{2})$, $(1, 2, 2)$ and vertex $(0, 2, 2)$, and six octahedra with quadrilateral faces, a typical octahedron having vertices $(0, 0, 2)$, $(0, 0, 3)$, $(\frac{1}{2}, \frac{1}{2}, 2\frac{1}{2})$, $(-\frac{1}{2}, \frac{1}{2}, 2\frac{1}{2})$, $(\frac{1}{2}, -\frac{1}{2}, 2\frac{1}{2})$, $(-\frac{1}{2}, -\frac{1}{2}, 2\frac{1}{2})$, $(\frac{2}{3}, 0, 2\frac{2}{3})$, $(0, \frac{2}{3}, 2\frac{2}{3})$, $(-\frac{2}{3}, 0, 2\frac{2}{3})$, $(0, -\frac{2}{3}, 2\frac{2}{3})$.

Diagrams have been given for the outer boundaries of the first four of these zones by Nicholas (1951).

§ 3. DENSITY OF STATES OF THE SEPARATE ZONES

If the electrons are considered to be perfectly free the number of states per unit volume (including the factor 2 to take account of both spin states) which lie within a single Brillouin zone is given by

$$\frac{m}{2a^2h^2k'}\iint dS,$$

where the integration is taken over the part of the surface of the sphere of radius k' which lies within this zone, k' being defined by equation (2). The above quantity is the density of states, $N(E)$, for the single zone.

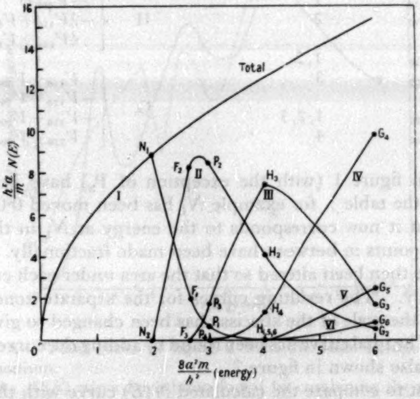


Figure 1. Density of states curves for the separate Brillouin zones of a body-centred cubic lattice, assuming the electrons to be entirely free. $N(E)\,h^2a/m$ is plotted against $8a^2mE/h^2$.

The integrals for the first six zones have been evaluated exactly for continuous values of k' up to $\sqrt{6}$. Since for free electrons, E is proportional to k'^2, the graph of $N(E)$ against k'^2 is proportional to the true density of states curve. These graphs are shown in figure 1, and it can be seen that there are discontinuities at the values of k'^2 equal to 2, $2\frac{2}{3}$, 3, 4 and 6. These points correspond to the start of new zones. In figure 1 N denotes the point corresponding to the point (1, 1, 0), in the Brillouin zones. In the same way F, P, H and G denote the points $(\frac{2}{3}, \frac{2}{3}, \frac{4}{3})$, (1, 1, 1), (0, 0, 2) and (1, 1, 2) in the zones. In figure 1 suffixes are added to the letters to indicate to which zone they belong.

§ 4. Modification due to Lattice Potential

Values for the discontinuities in the energy at the points N, F, P, H and G can be found in terms of the Fourier coefficients of potential energy, by using the approximation of nearly free electrons. Approximate values of the Fourier coefficients V_n have been found by fitting the energy gaps at N and H to the data for sodium given by Howarth and Jones (1952), the resulting values for V_{110} and V_{200} being 0·024 and 0·039 rydberg unit respectively. In order to find V_{211} a spherical potential of the form $(1/r)\exp(-\alpha r)$ was assumed in the atomic cell, and the ratio of V_{110} to V_{211} obtained. The value adopted for V_{211} was 0·054 rydberg unit.

The different energies associated with each of the points N, F, P, etc. can be correlated with the separate Brillouin zones, the lowest energy corresponding to the lowest zone, the second lowest energy to the second lowest zone and so on.

The values of the energy relative to the energy of a free electron at the same point of the zone are given in the table.

Point	Energy	Zone	Point	Energy	Zone
N	$-V_{110}$	1	H	$-V_{200}$	1, 2, 3
	V_{110}	2		$-2V_{110}+V_{200}$	4, 5
F	$-V_{110}$	1, 2		$4V_{110}+V_{200}$	6
	$2V_{110}$	3	G	$V_{110}-V_{200}-V_{211}$	3
P	$-V_{110}$	1, 2, 3		$-V_{110}+V_{200}-V_{211}$	4
	$3V_{110}$	4		$-V_{110}-V_{200}+V_{211}$	5
				$V_{110}+V_{200}+V_{211}$	6

The points in figure 1 (with the exception of P_3) have been displaced in accordance with the table ; for example N_1 has been moved 0·024 rydberg unit backwards so that it now corresponds to the energy at N_1 in the actual metal. Displacement of points in between have been made fractionally. The ordinates of the curves have then been altered so that the area under each curve is the same as it was originally. The resulting curves for the separate zones are shown in figure 2 in which the scale of the abscissa has been changed to give the energy in electron volts. The total curve has been found by adding the curves of the separate zones and this is also shown in figure 2.

It is of interest to compare the calculated $N(E)$ curve with the observed soft X-ray emission and absorption spectra. A curve relating to lithium may be constructed from the data given by O'Bryan and Skinner (1934) for the soft X-ray emission, and by Skinner and Johnston (1937) for the soft X-ray absorption. The experimental curves giving the intensities of the emission, and the absorption, against frequency may be converted into plots against energy, and the vertical scales of intensity can be chosen so that the first peaks on both curves have the same height. Such a curve is shown by the full line in figure 3. There is, of course, inevitably some uncertainty in the fitting of the emission and the absorption data together. A theoretical curve which may be compared with this can be obtained

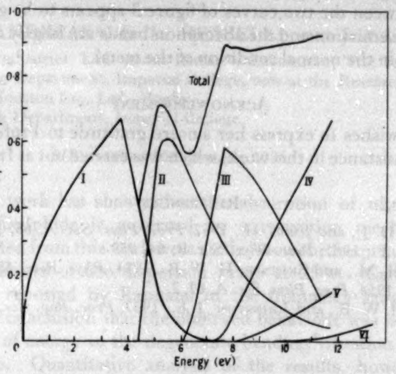


Figure 2. Modified density of states for the separate Brillouin zones in the case of sodium.

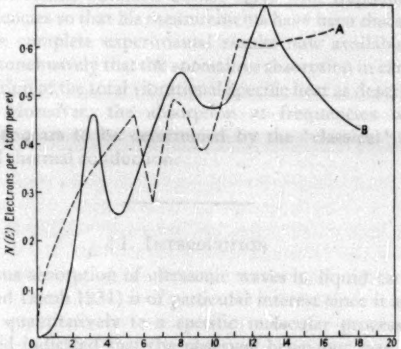


Figure 3. Curve A: Total curve of figure 2 with the scale change to ev for lithium instead of sodium.

Curve B: $N(E)$ curve for lithium found by combining the soft X-ray emission spectra with the soft X-ray absorption spectra.

by changing the value of the lattice constant in the factor which relates E and k'^2. Thus from the total density of states curve of figure 2 a similar curve can be obtained with an energy scale appropriate to metallic lithium. This is shown as the broken curve in figure 3. It is known from the calculations made by Schiff (1954) that the energy gap at N_1 is approximately two electron volts larger in lithium than in sodium. Thus the first peak of the dotted curve should be displaced approximately one electron volt to the left. It is not possible to say from existing calculations how the higher energy peaks should be displaced. However the general resemblance between the two curves of figure 3 appears to be good evidence that the shapes of the emission and the absorption bands are largely determined by the density of states in the normal condition of the metal.

ACKNOWLEDGMENT

The author wishes to express her sincere gratitude to Professor H. Jones for his advice and assistance in this work, which was carried out at Imperial College.

REFERENCES

HOWARTH, D. J., and JONES, H., 1952, *Proc. Phys. Soc.* A, **65,** 355.
NICOLAS, J. F., 1951, *Proc. Phys. Soc.* B, **64,** 953.
O'BRYAN, H. M., and SKINNER, H. W. B., 1934, *Phys. Rev.*, **45,** 370.
SCHIFF, B., 1954, *Proc. Phys. Soc.* A, **67,** 2.
SKINNER, H. W. B., and JOHNSTON, J. E., 1937, *Proc. Roy. Soc.* A, **161,** 420.

Ultrasonic Relaxation and the Vibrational Specific Heat of Carbon Disulphide

By J. H. ANDREAE†, E. L. HEASELL‡ and J. LAMB§

† Imperial Chemical Industries, Ltd. Akers Research Laboratory, The Frythe, Welwyn.
‡ Electrical Engineering Department, Imperial College, now at the Research Laboratories, British Thomson-Houston Co., Ltd., Rugby.
§ Electrical Engineering Department, Imperial College.

MS. received 10th January 1956

Abstract. Previous work has shown that the absorption of ultrasonic waves in liquid carbon disulphide is governed by a relaxation mechanism. The measurements reported from this work were made below the frequency of 70 Mc/s, which is near the centre of the dispersion range at 25°c. Combining these results with others reported by Rapuano in the frequency range 75 Mc/s to 195 Mc/s, led to the conclusion that the observed behaviour was associated with the time of transfer of energy to the degenerate bending vibration of the carbon disulphide molecule. Quantitative analysis of the results, however, was not regarded as satisfactory.

Measurements of the absorption of ultrasonic waves in carbon disulphide have now been made by the authors at frequencies up to 190 Mc/s and at different temperatures. The results obtained do not agree with Rapuano's particularly at the higher frequencies so that his measurements have been discarded.

Analysis of the complete experimental results now available at both 25°c and −63°c shows conclusively that the anomalous absorption in carbon disulphide is due to the relaxation of the total vibrational specific heat as described by a single relaxation time. Moreover, the absorption at frequencies well above the relaxation region appears to be determined by the 'classical' mechanisms of shear viscosity and thermal conduction.

§ 1. Introduction

The anomalous absorption of ultrasonic waves in liquid carbon disulphide (Andreae and Lamb 1951) is of particular interest since it appears possible to relate it quantitatively to a specific molecular process. The results previously obtained indicated that the observed behaviour was associated with the time required to establish equilibrium between the translational energy and the energy contained in the degenerate bending vibration of the CS_2 molecule. In order to make a complete analysis of the behaviour it is necessary to have available experimental results covering a substantial part of the dispersion range. The previous results of Andreae and Lamb extended only up to a frequency of 70 Mc/s which is near the centre of the dispersion region at 25°c. Use was made, however, of additional results reported by Rapuano (1950) in the frequency range 75 Mc/s to 195 Mc/s. The quantitative analysis of the combined data led to complications (Andreae and Lamb 1954) which become superfluous in the light of new measurements. With improved equipment the present authors have now measured the ultrasonic absorption in carbon disulphide up to 190 Mc/s. The results obtained are reported in this paper and since they do not agree with Rapuano's, particularly at the higher frequencies, the latter have been discarded. In consequence a satisfactory analysis can now be made, providing the first instance in which an ultrasonic relaxation in a liquid has been quantitatively related to a specific molecular process in an unambiguous manner.

§ 2. EXPERIMENTAL SYSTEMS

All the measurements of the absorption of ultrasonic waves in carbon disulphide have been made by what is generally termed the 'pulse method' after Pellam and Galt (1946) and Pinkerton (1949). Three different sets of experimental apparatus have been employed over the frequency range 20 Mc/s to 190 Mc/s. Since an adequate description of the technique is available (Pinkerton 1949), it suffices to give a brief outline of the experimental arrangements and to indicate the particular features of each apparatus which differ from previous descriptions.

Separate X-cut quartz crystals are employed to generate and to detect the ultrasonic vibrations. Pulses of radio frequency voltage of approximately 1 kv amplitude, peak to peak, and of 5 microseconds duration are applied to the transmitting quartz crystal at a repetition frequency of approximately 500 per second. The transducer is acoustically connected by a thin film of oil to a length of fused quartz rod which is optically polished on its end faces and serves as a delay line. The time taken for an ultrasonic wave to make a double journey in the rod exceeds the duration of the pulse ($5\,\mu$sec). The upper end of the rod is in contact with the liquid which is contained in a suitable vessel attached to the fused quartz rod by means of a ground joint.

A similar fused quartz rod with a receiving transducer attached to its upper end is supported on a movable carriage. The lower end of this rod dips into the liquid. Changes of the ultrasonic path length are recorded on a dial gauge.

The fused quartz rods are enclosed in a chamber which is thermostatically controlled to within $\pm 0{\cdot}1^\circ$c, the temperature being measured by means of a thermocouple which is attached to the base of the movable rod.

The pulses of radio frequency generated by the receiving transducer are amplified by a high-gain receiver and are finally displayed, after rectification, as video pulses on the screen of a cathode-ray tube.

Three sets of experimental arrangements have been used: these differ mainly in the mechanical construction employed to produce known changes in the ultrasonic path whilst maintaining parallelism between the end faces of the two quartz rods. The systems differ electrically in that in A a piston attenuator is connected in series with the output from the receiving crystal while in B the attenuator is placed in series with the electrical feed to the transmitting crystal. The settings of the attenuators are adjusted to compensate for changes in the ultrasonic path, the comparison being effected by visual observation of the height of the pulse envelope on the screen of the cathode-ray tube.

In the third system C, use is made of a pulsed comparison oscillator operating at the same frequency as the oscillator which energizes the transmitting crystal. The pulse which is applied to the comparison oscillator is derived from the main pulse generator and occurs at a suitable time interval after the main transmitter pulse. The output from this oscillator is fed through an attenuator and then receives the same amplification as the signal which passes through the ultrasonic system. The variable time-delay is adjusted so that the comparison pulse appears on the screen of the cathode-ray tube alongside the first received pulse from the receiving quartz crystal. The attenuator setting is again adjusted to equalize the height of the two pulses when changes are made in the ultrasonic path. In this method of operation the gain of the amplifier is adjusted to give approximately the same output from the receiver under all conditions.

The two piston attenuators which have been employed in this work are of circular cross section and operate in the E_{01} mode. They have different but accurately known bores and have been checked against other attenuators operating in the H_{11} (circular) and H_{01} (rectangular) modes. In all cases the calculated attenuation constants were found to apply in practice to within the limiting accuracy ($\pm\frac{1}{2}\%$) of the method of comparison.

Different pairs of X-cut quartz crystals have been used, the choice of the fundamental frequency of thickness vibration being determined by the desired spacing of the harmonic frequencies. Fundamental resonance frequencies in the range 7·5 Mc/s to 20 Mc/s have been employed.

The three sets of apparatus cover approximately the same frequency range and the agreement between the values measured on these different systems is taken as sufficient evidence for rejecting the conflicting data given by Rapuano.

Analar grade carbon disulphide manufactured by Messrs. Hopkin and Williams Ltd. has been used for most of the work. Measurement of samples immediately after careful distillation revealed no difference in the absorption values.

No published results could be found for the velocity of sound in carbon disulphide at -63°c and at frequencies below the relaxation range. These values were measured by means of an interferometer which was operated at a frequency of 1·75 Mc/s and was of conventional design employing a fixed reflector and a movable crystal transducer.

§ 3. THE ULTRASONIC DATA

An attempt has been made in the experimental work to cover as much as possible of the dispersion range which is centred around a frequency of 78 Mc/s at 25°c. Previous absorption measurements have been made by Huddart (1950) at frequencies of 2, 6 and 10 Mc/s and over the temperature range 0°c to 35°c. His values at 25°c are given in table 1 which contains the results of the present work. The values quoted in this table are in each case the average of a number of observations taken under identical conditions. Although measurements have been made over a temperature range the observations at 25°c and -63°c, which are given here, were taken with particular care. The lowest temperature (-63°c) was determined by practical considerations, the aim being to obtain accurate results at two widely different temperatures, for the purpose of theoretical analysis.

The ultrasonic velocity in CS_2 has been measured to a high order of accuracy by Freyer, Hubbard and Andrews (1929) at a frequency of 414 kc/s. These workers found that the velocity is proportional to temperature over the temperature range 0°c to 40°c, the value at 25°c being $c = 1142\,\mathrm{m\,sec^{-1}}$. Extrapolation of the linear law to temperature below 0°c has been verified by the authors, the measured value at -63°c being $1424\,\mathrm{m\,sec^{-1}}$.

Table 1. Experimental Results for the Absorption in Carbon Disulphide at Different Frequencies

α = absorption coefficient relating to the amplitude of excess pressure

f (Mc/s)		2†	6†	10†	23	29·6	37·4	48·5	66·7
$\frac{\alpha}{f^2}\times 10^{17}$ (sec² cm⁻¹)	25°c	5830	5910	5650	5390	5100	4650	4120	3240
	−63°c	—	—	—	3920	—	2430	—	1090

f (Mc/s)		81·2	103·5	106·2	107·4	111	148·2	168	189·3
$\frac{\alpha}{f^2}\times 10^{17}$ (sec² cm⁻¹)	25°c	2800	2060	1960	1870	1955	1200	970	760
	−63°c	795	—	—	—	466	—	—	—

† Huddart (1950)

§ 4. PHYSICAL PROPERTIES OF CARBON DISULPHIDE

Carbon disulphide has a molecular weight $M = 76{\cdot}13$: at atmospheric pressure it freezes at −111°c and boils at +46·3°c. The critical point is at a temperature of 273°c and a pressure of 76 atmospheres.

A summary is given in table 2 of the various physical constants required in the following analysis of § 5.

Table 2. Physical Properties

Temperature	T (°c)	+25	−63
Density*	ρ (g cm^{-3})	1·256	1·385
Thermal expansion coefficient*	θ (deg)$^{-1}\times 10^3$	1·191	1·041
Ultrasonic velocity at 414 kc/s‖	c (m sec^{-1})	1142	1424
Shear viscosity	η (cpoise)	0·355†	0·8§
Specific heat at constant pressure †	C_p (cal mole^{-1} deg^{-1})	18·17	17·97
Ratio of specific heats ‡	$\gamma = C_p/C_v$	1·551	1·465
Specific heat at constant volume	C_v (cal mole^{-1} deg^{-1})	11·71	12·25

* Timmermans (1950).

‖ Freyer, Hubbard and Andrews (1929) and authors' measurement at −63°c.

§ From extrapolation formula given in *International Critical Tables*, 1929, **7**, 213.

† Brown and Manov (1937).

‡ Calculated from the formula $\gamma - 1 = Mc^2\theta^2 T/C_p$.

§ 5. ANALYSIS OF RESULTS

The pronounced decrease in the value of α/f^2 for carbon disulphide shown by the results recorded in table 1 occurs at frequencies above 2 Mc/s. As a preliminary to analysing these results it is necessary to consider the possibility or otherwise of other relaxation mechanisms occurring at frequencies below 2 Mc/s. The adiabatic compressibility of CS_2 has been measured by Tyrer (1913, 1914) at very low frequencies. Tyrer's method does not involve a precise frequency, but from the information given in his paper this would appear to be in the region of one cycle per second. Very good agreement (closer than 1%) is found between the adiabatic compressibility measurements of Tyrer and the values derived from the ultrasonic velocity data by Freyer, Hubbard and Andrews (1929), the latter corresponding to a frequency of 414 kc/s. It is concluded therefore that no significant dispersion occurs between 414 kc/s and 1 c/s and it is very unlikely that any such effect would occur at frequencies below 1 c/s. Consequently the value of the 'static' specific heat $(C_p)_s$ will be taken to apply at frequencies just below the present relaxation range (<2 Mc/s). The corresponding value of $(C_V)_s$ is then calculated using the formula $\gamma_s - 1 = Mc_s{}^2\theta_s{}^2T/(C_p)_s$, where the suffix s denotes a low-frequency or 'static' value.

Preliminary analysis of the results of table 1 indicated that the observed behaviour is not due to the time of transfer of energy to the degenerate bending vibration only but rather to a relaxation of the total vibrational specific heat.

The fundamental frequencies ν of the normal modes of vibration of the CS_2 molecule are given by Herzberg (1945) and these together with the calculated values of the contributions to the vibrational specific heat are recorded in table 3.

Table 3. Vibrational Specific Heat

Vibrational frequency ν (cm^{-1})		657	397	1523
Degeneracy n		1	2	1
Energy quantum ($h\nu$) (kcal)		1·877	1·134	4·352
Einstein specific heat (cal mole^{-1} deg^{-1}) †	+25°c	0·9130	2·951	0·0690
	−63°c	0·4569	2·220	0·0064
Total vibrational specific heat ΔC (cal mole^{-1} deg^{-1})	+25°c		3·933	
	−63°c		2·683	

† $(=nRx^2e^{-x}/(1-e^{-x})^2$, where $x = h\nu/kT)$

In the present case of a vibrational relaxation the elevation of a molecule to a higher internal energy state is likely to take place at constant volume. Herzfeld (1941), Davies (1954), Davies and Lamb (1956) Andreae and Lamb (1956) and others have shown that in these circumstances the specific heats at constant pressure C_p and constant volume C_V each contain the relaxing component $\Delta C/(1+j\omega\tau)$, where τ is a relaxation time and ω is 2π times the frequency. ΔC is the contribution made by the molecular vibrations to the static specific heats $(C_p)_s$ and $(C_V)_s$ (i.e. the specific heats as $\omega \to 0$). These quantities may be related to the absorption α and phase velocity c of ultrasonic waves by the formulae

$$\frac{\alpha c}{f^2} = \frac{A}{[1+(f/f_c)^2]} + B \qquad \ldots\ldots(1)$$

$$\left(\frac{c}{c_s}\right)^2 = 1 + \frac{\{(C_p)_s - (C_V)_s\}\Delta C}{(C_p)_s\{(C_V)_s - \Delta C\}}\,\frac{1}{1+(f_c/f)^2} \qquad \ldots\ldots(2)$$

where

$$f_c = \frac{1}{2\pi\tau}\left[\frac{(C_p)_s(C_V)_s}{\{(C_p)_s - \Delta C\}\{(C_V)_s - \Delta C\}}\right]^{1/2} \qquad \ldots\ldots(3)$$

and c_s is the low frequency value of the phase velocity c. The derivation of these formulae, as given by the above-mentioned authors, proceeds from the assumption that a process relaxes with a single relaxation time τ and that no other process has a relaxation time of the same order of magnitude as τ. Also it is assumed that $(\alpha c/\omega)^2$ is much less than unity.

It follows from formula (1) that the ultrasonic absorption per wavelength, $\mu = \alpha c/f$, is composed of two parts: the one equal to Bf represents the combined effects of shear viscosity, heat conduction and all processes having much smaller relaxation times than τ; the other part,

$$\mu' = \frac{Af}{1+(f/f_c)^2} = \mu_m'\,\frac{2(f/f_c)}{1+(f/f_c)^2}, \qquad \ldots\ldots(4)$$

is the absorption per wavelength due to the single relaxing process and it attains a maximum value

$$\mu_m' = \frac{\pi}{2}\frac{\{(C_p)_s-(C_v)_s\}\Delta C}{[(C_p)_s\{(C_p)_s-\Delta C\}(C_v)_s\{(C_v)_s-\Delta C\}]^{1/2}}, \qquad \ldots\ldots(5)$$

at the frequency $f=f_c$.

Ultrasonic measurements have provided the value of c_s and the variations of α/f^2, (not $\alpha c/f^2$), with frequency. Since the velocity dispersion cannot be neglected in the case of carbon disulphide, comparison of the measured α/f^2 values with the calculated values can only be effected by a series of successive approximations. First, it is assumed that a relaxation of the total vibrational specific heat is responsible for the observed behaviour and, as will be evident later, this is found to be the case. Secondly, an assumed value for f_c is substituted in equation (2) to give the velocity c, which is then used in the evaluation of α/f^2 from equation (1). In making this evaluation the parameter A is chosen to give the best fit with the experimental values of α/f^2. The process is then repeated until the calculated and measured values agree within experimental accuracy.

Table 4. Testing the Relaxation Formulae (1) and (2) for CS_2 at 25°C

$\Delta C=3{\cdot}933$ (cal mole^{-1} deg^{-1}), the value of the total vibrational specific heat

$A=6700\times10^{-12}$ sec; $f_c=78$ Mc/s.

$$\left(\frac{\alpha}{f^2}\right)_{\text{calc}}=\frac{A/c}{1+(f/f_c)^2}:\quad\left(\frac{c}{c_s}\right)^2=1+\frac{\{(C_p)_s-(C_v)_s\}\Delta C}{(C_p)_s\{(C_v)_s-\Delta C\}}\,\frac{1}{1+(f_c/f)^2}$$

Experimental arrangement				A	A	A	A	A	A
f (Mc/s)		<6	10	23	29·6	37·4	48·5	66·7	81·2
c (m sec^{-1})		1142	1144	1150	1155	1161	1170	1184	1194
$\frac{\alpha}{f^2}\times10^{17}$ (sec^2 cm^{-1})	exptl	5870*	5650†	5390	5100	4650	4120	3240	2800
	calc	5829	5763	5359	5071	4692	4129	3267	2692
% excess of experimental over calculated values		+0·7	−2·0	+0·6	+0·6	−0·9	−0·2	−0·8	+3·8

Experimental arrangement		C	C	C	A	C	B	C
f (Mc/s)		103·5	106·2	107·4	111	148·2	168·0	189·2
c (m sec^{-1})		1206	1207	1207	1208	1219	1224	1227
$\frac{\alpha}{f^2}\times10^{17}$ (sec^2 cm^{-1})	exptl	2060	1960	1870	1955	1200	970	776
	calc	2013	1945	1916	1832	1191	971	793
% excess of experimental over calculated values		+2·3	+0·8	−2·4	+6·7	+0·7	−0·1	−2·2

μ_m' (calc) (from eqn (5))$=0{\cdot}260$: μ_m' (exptl)$=\tfrac{1}{2}Af_c=0{\cdot}262$

* This is the average of Huddart's (1950) values at 2 and 6 Mc/s. † Huddart (1950).

Finally, the validity of the whole argument is examined by comparing the value of $\mu_m'(=\tfrac{1}{2}Af_c)$, selected to give the best fit with the experimental results, with the value calculated from equation (5) on the basis that ΔC is the total vibrational specific heat.

It transpires that good agreement between theory and experiment is obtained at 25°C when the term B in equation (1) is of the order of the contribution due to shear viscosity and thermal conduction. This amount is less than 1% of the measured α/f^2 value at the highest frequency (189·2 Mc/s) and has therefore been neglected. The results of these calculations at 25°c are given in table 4 from which it can be seen that there is excellent agreement between the calculated and experimental values of μ_m'. Agreement between the calculated and experimental value of α/f^2 is within the experimental accuracy, estimated to be to ± 3%, except at 81·2 Mc/s and 111 Mc/s. However, absorption values at these two frequencies were taken with the first experimental arrangement A, the mechanical design of which was less satisfactory than for systems B and C in relation to measurements at the higher frequencies of the present range. The extent of the agreement shown in table 4 is regarded as satisfactory, more reliance being placed on measurements taken with systems B and C above 100 Mc/s.

At -63°c measurements were made only with the experimental system A, but these were sufficient to confirm the agreement found at 25°c between the experimental and calculated values of μ_m'. The results are given in table 5.

Table 5. Testing the Relaxation Formulae (1) and (2) for CS_2 at -63°c

$\Delta C = 2{\cdot}683$ (cal mole^{-1} deg^{-1}) (table 3)

$A = 8650 \times 10^{-12}$ sec; $f_c = 31$ Mc/s; $B = 65 \times 10^{-12}$ sec †

	Experimental arrangement A				
f (Mc/s)	23	37·4	66·7	81·2	111
c (m sec^{-1})	1446	1461	1475	1478	1481
$\dfrac{\alpha}{f^2} \times 10^{17}$ (sec^2 cm^{-1}) exptl	3920	2430	1090	795	466
$\dfrac{\alpha}{f^2} \times 10^{17}$ (sec^2 cm^{-1}) calc	3902	2455	1086	788	466
% excess of experimental over calculated values	+0·5	−1·0	+0·5	+0·9	—

μ_m' (calc) (from eqn (5)) $= 0{\cdot}133$: μ_m' (exptl) $= \tfrac{1}{2} A f_c = 0{\cdot}134$.

† This value of B has been chosen to give the best fit to the experimental points. It is somewhat larger than the classical contribution due to shear viscosity ($7{\cdot}5 \times 10^{-12}$ sec) obtained by using an extrapolation formula for the shear viscosity at -63°c (see table 2). In view of the limited experimental accuracy obtained with the experimental arrangement A at the higher frequencies, this discrepancy is regarded as having no significance.

§ 6. CONCLUSIONS

It is concluded from the analysis of § 5 that the anomalous absorption in carbon disulphide is due to a relaxation of the total vibrational specific heat as described by a single relaxation time. Very good agreement is found at both 25°c and -63°c between the two values of the maximum absorption per wavelength which are calculated from the total vibrational specific heat and from the measured ultrasonic absorption, respectively. It appears that at frequencies above the relaxation region the absorption falls to near the 'classical' value attributed to shear viscosity and thermal conduction.

ACKNOWLEDGMENT

The work reported in this paper was supported by a grant from the Department of Scientific and Industrial Research. The authors gratefully acknowledge the assistance provided by this grant.

References

ANDREAE, J. H., and LAMB, J., 1951, *Proc. Phys. Soc.* B, **64,** 1021 ; 1954, *Proc. Roy. Soc.* A, **226,** 51 ; 1956, *Proc. Phys. Soc.* B, **69,** in the press.

BROWN, O. L. I., and MANOV, G. G., 1937, *J. Amer. Chem. Soc.*, **59,** 500.

DAVIES, R. O., 1954, *Proc. Roy. Soc.* A, **226,** 24.

DAVIES, R. O., and LAMB, J., 1956, *Proc. Phys. Soc.*, **69,** 293.

FREYER, E. B., HUBBARD, J. C., and ANDREWS, D. H., 1929, *J. Amer. Chem. Soc.*, **51,** 759.

HERZBERG, G., 1947, *Molecular Spectra and Molecular Structure* II. *Infra-red and Raman Spectra of Polyatomic Molecules* (New York : Van Nostrand).

HERZFELD, K. F., 1941, *J. Acoust. Soc. Amer.*, **13,** 33.

HUDDART, D. H. A., 1950, *M.Sc. Thesis*, University of London.

PELLAM, J. R., and GALT, J. K., 1946, *J. Chem. Phys.*, **14,** 608.

PINKERTON, J. M. M., 1949, *Proc. Phys. Soc.* B, **62,** 286.

RAPUANO, R. A., 1950, *Res. Lab. Electronics Tech. Rep. No.* 151, Massachusetts Institute of Technology.

TIMMERMANS, J., 1950, *Physico-chemical Constants of Pure Orangic Compounds* (Amsterdam: Elsevier).

TYRER, D., 1913, *J. Chem. Soc.*, **103,** 1675; 1914, *Ibid.*, **105,** 2534.

Paramagnetic Resonance in Chromic Methylamine Alum

By J. M. BAKER
Clarendon Laboratory, Oxford

Communicated by B. Bleaney; *MS. received 7th November* 1955, *and in amended form* 13*th January* 1956

Abstract. The paramagnetic resonance spectrum of chromic methylamine alum has been re-examined because of an 11% discrepancy in the value of the Stark splitting between previous paramagnetic resonance and other low temperature measurements. The crystalline electric field acting on the chromic ion, which has trigonal symmetry at high temperatures, is found to undergo a change below about 160°K to rhombic symmetry. The Stark splitting is then $0{\cdot}178 \pm 0{\cdot}002\,\mathrm{cm}^{-1}$; this is a little higher than the previous resonance value, reducing the discrepancy to 5%. Measurements of line width indicate that this discrepancy is probably not due to exchange interaction.

---

§ 1. INTRODUCTION

CHROMIC methylamine alum, $Cr(CH_3NH_3)(SO_4)_2\ 12H_2O$, forms crystals with cubic symmetry, space group T_h^6, with four molecules per unit cell. The alum has a β-type structure (Lipson and Beevers 1935, Lipson 1935) for which X-ray analysis shows the chromic ion to be at the centre of a regular octahedron of water molecules. These give a strong crystalline electric field of cubic symmetry, which lifts the sevenfold orbital degeneracy of the ground state of the free chromic ion $^4F_{3/2}$, leaving an orbital singlet with fourfold spin degeneracy lowest.

Paramagnetic resonance measurements of Bagguley and Griffiths (1950), made at room temperature, show that this quadruplet is split into two spin doublets by a small axial component of the crystal field, which is produced by a trigonal distortion of the octahedron of water molecules, the trigonal axis being parallel to a body diagonal of the cubic structure. Each of the four chromic ions in the unit cell has the axis of its trigonal distortion parallel to a different body diagonal. The ground state may be described by the spin Hamiltonian:

$$\mathscr{H} = g\beta\mathbf{H}\cdot\mathbf{S} + D\{S_z^2 - \tfrac{1}{3}S(S+1)\} \qquad \ldots\ldots(1)$$

where H is the external magnetic field, β is the Bohr magneton, $S = 3/2$ and the value of g is about 1·97 (there is also a hyperfine structure due to ^{53}Cr, whose natural abundance is 9·5%, but as this is not resolved in this work it may be neglected). The z axis for each ion is that of the trigonal distortion of its crystal field. The separation of the spin doublets in zero field is $\delta = 2D$, and Bagguley and Griffiths found this to be $0{\cdot}165 \pm 0{\cdot}005\,\mathrm{cm}^{-1}$ at room temperature.

Bleaney (1950), also using paramagnetic resonance, found that $2D$ increases to $0{\cdot}170 \pm 0{\cdot}003\,\mathrm{cm}^{-1}$ at 90°K and shows no further change when the temperature is reduced to 20°K. Cooke and Duffus (unpublished), using paramagnetic resonance, found the same value at 4°K. The low temperature measurements were made only with the external magnetic field parallel to a body diagonal, and the value of δ was taken as $2D$, assuming the crystal field to have the same symmetry as at room temperature.

The separation of the spin doublet δ has been measured by other less direct methods with slightly different results. Gardner and Kurti (1954) and Hudson and McLane (1954), both using adiabatic demagnetization methods, obtained respectively values of $0{\cdot}188 \pm 0{\cdot}007\,\text{cm}^{-1}$ and $0{\cdot}187 \pm 0{\cdot}002\,\text{cm}^{-1}$. Paramagnetic relaxation measurements of Cooke and Wolf (1953, unpublished) confirm the value of the magnetic specific heat obtained by Gardner and Kurti. There is thus a discrepancy between these results, which agree remarkably well with one another, and those of paramagnetic resonance, which is well outside the experimental error. The paramagnetic resonance spectrum of the substance has been studied again in an attempt to resolve this difference, and this paper is concerned with the results of this investigation.

§ 2. EXPERIMENTAL DETAILS

In strong magnetic fields the spectrum (calculated from (1)) of each ion consists of a 'central' transition ($M=\frac{1}{2}$ to $M=-\frac{1}{2}$) which occurs at $h\nu=g\beta H$ (apart from second order shifts), and two 'outer' transitions ($M=\pm\frac{3}{2}$ to $M=\pm\frac{1}{2}$) which are separated by $2D(3\cos^2\theta-1)$ when the field H makes an angle θ with the trigonal axis. The measurements described below have been made in two crystal planes; (111), which is easily found because it is parallel to one of the prominent crystal faces, and $(1\bar{1}0)$, which contains the three directions [001], [110] and [111] where the spectrum should be particularly simple (see Bagguley and Griffiths). This plane is obtained by placing two adjacent {111} faces of the crystal flush with the sides of a V-shaped trough of the correct angle, cut in distrene. Using these techniques it is possible to adjust the crystal so that the angle between one of these planes and the plane of rotation of H is not greater than 1°. This is confirmed by the fact that the spectrum of chromic potassium selenate, when mounted in either of these planes, shows the expected behaviour to within the experimental error.

The standard technique of modulated external field and video frequency detection was used for all measurements in strong fields (central line at approximately 8 kilo-oersteds). The measurements at low fields employed in some cases the older galvanometer technique described by Bagguley, Bleaney, Griffiths, Penrose and Plumpton (1948) where the power absorbed by the crystal is plotted point by point as a function of the external field. The field for the low field experiments was produced by a solenoid which was calibrated in terms of current, using proton magnetic resonance at high currents. Strong fields were measured directly by proton resonance.

§ 3. RESULTS

3.1. *Measurements at High Fields*

At temperatures of 90°K and below the angular dependence of the spectrum upon orientation of the external field was not that expected from the above Hamiltonian. As the resolution in the undiluted salt was poor the line width was reduced from ∼100 oersteds to ∼10 oersteds by using a diluted crystal, i.e. a mixed crystal of 99% aluminium methylamine alum and 1% of the isomorphous chromic salt. This showed a similar angular dependence which was easier to elucidate because of the better resolution. The departure from trigonal symmetry is shown very clearly in the {111} plane (figure 1 (*a*)). Since the magnetic field is always normal to the {111} axis, which (see § 1) is parallel to the *z* axis of one ion, the spectrum of this ion (marked by the continuous lines) should remain fixed, while the others should show a 60° repetition. Also, whereas with trigonal symmetry about {111} axes all {110} planes are equivalent, in this case there are two types of spectra from these planes (figure 1 (*b*) and (*c*)).

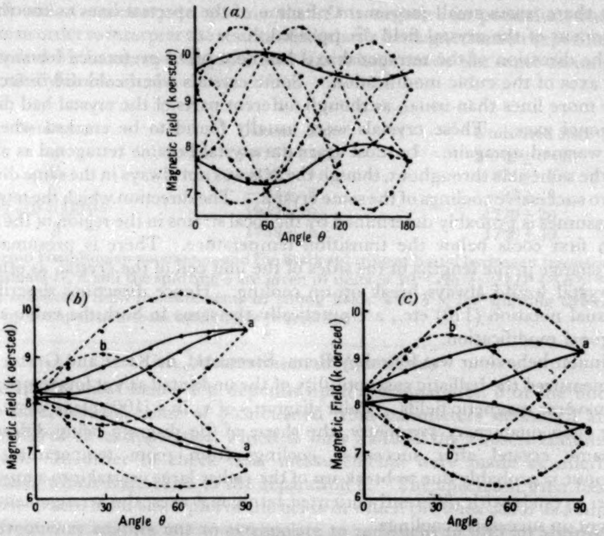


Figure 1. The fields at which the outer lines occur as a function of the angular orientation of the field in various planes. (*a*) {111} plane : if the crystal field had trigonal symmetry the heavy lines would be parallel to the θ axis and the whole spectrum would have a 60° periodicity ; (*b*) and (*c*) (110) and (011) planes : if the crystal field had trigonal symmetry these would be identical and the transitions for the two ions marked with full lines would be coincident, with the separation aa equal to bb.

The spectra in the {100} planes (not illustrated) and {110} planes show reflection symmetry in the cube faces so that the overall symmetry cannot be lower than orthorhombic. The fact that there are only two types of {110} plane (figure 1(*b*) and (*c*)) restricts the symmetry to not lower than tetragonal, the (110) and $(1\bar{1}0)$ planes (taking *c* as the tetragonal axis) having a spectrum like figure 1 (*b*), and the other four {110} planes being like figure 1 (*c*).

The observed angular dependence in all planes can be accounted for in terms of a rhombic distortion of the crystal field, which modifies the Hamiltonian to

$$\mathscr{H} = g\beta\mathbf{H}\cdot\mathbf{S} + D[S_z^2 - \tfrac{1}{3}S(S+1)] + E(S_x^2 - S_y^2) \qquad \text{......(2)}$$

the value of *E* being about 10% that of *D*. To satisfy the symmetry consideration of the last paragraph the (110) and $(1\bar{1}0)$ planes each contain the *z* axis of two ions.

A rhombic distortion of the crystal field is incompatible with the space group T_h with four ions in the unit cell. As the spectrum shows only four ions, the space group at low temperatures must be one of lower symmetry, probably D_{4h}, the tetragonal crystal axis being parallel to one of the cubic axes of the T_h modification. Thus, there must be a transition in crystal structure between room temperature and 90°K. This transition has been observed in both diluted and undiluted crystals when they were allowed to warm up from 90°K; at about 160°K there was a small movement of some of the spectral lines as the rhombic component of the crystal field disappeared.

The direction of the tetragonal axis has no *a priori* preference for any of the three axes of the cubic modification. Some crystals when cold did in fact show many more lines than usual, as though different parts of the crystal had different tetragonal axes. These crystals were usually found to be cracked when they were warmed up again. In most cases the crystal became tetragonal as a whole with the same axis throughout, though the axis was not always in the same direction for two successive coolings of the same crystal. The direction which the tetragonal axis assumes is probably determined by the local strains in the region of the crystal which first cools below the transition temperature. There is presumably no great change in the lengths of the sides of the unit cell of the crystal, as otherwise the crystal would always break up on cooling. Hence directions described by the usual notation (110) etc., are practically the same in both the cubic and the tetragonal modification.

Similar behaviour was found by Beun, Steenland, de Klerk and Gorter (1955) who measured the ballistic susceptibility of the undiluted salt at low temperatures in transverse magnetic fields. Polar diagrams of $x_\perp$ in a {100} plane show binary rather than quaternary symmetry, the shape of the diagram being different for the same crystal after successive coolings from room temperature. This behaviour is probably due to break-up of the rather large crystal into a number of smaller crystals with differently oriented tetragonal axes, the arrangement being different on successive coolings.

The directions of the rhombic axes in the undiluted salt are given below in the form of direction cosines with respect to the orthogonal axes of the tetragonal crystal, the c axis being that of the tetragonal axis:

	a axis	*b* axis	*c* axis
x axis	$-0{\cdot}35 \pm 0{\cdot}05$	$-0{\cdot}35 \pm 0{\cdot}05$	$+0{\cdot}87 \pm 0{\cdot}03$
y axis	$+0{\cdot}71 \pm 0{\cdot}04$	$-0{\cdot}71 \pm 0{\cdot}04$	$0 \pm 0{\cdot}05$
z axis	$+0{\cdot}61 \pm 0{\cdot}04$	$+0{\cdot}61 \pm 0{\cdot}04$	$+0{\cdot}50 \pm 0{\cdot}04$

The directions for the other three ions are obtained by successive rotations of the system axes (xyz) about the c axis through angles of $\pi/2$. As it is established that the z axis lies in one of two planes through symmetry arguments, for whose validity the fact that all the figure 1 (c) type planes are identical is a fairly sensitive test, the largest error in the determination of the direction of the z axis is in deciding its orientation within the plane illustrated in figure 1 (b), which can be done to about 3°. This has been taken as the error for all the directions given above. Amongst other results of the transition from cubic to tetragonal symmetry the following may be noted: (a) the angle which the z axis makes with the 001 axis has changed from 55° to 60 ± 3°; (b) one of the rhombic axes (the y axis, assuming E to have the same sign as D) is along a face diagonal of the cubic modification.

The values of g, D and E were obtained in the usual way and are given in the table, together with the resultant separation of the doublets in zero field, $\delta = 2(D^2+3E^2)^{1/2}$. The value of g is identical with that ($1{\cdot}976 \pm 0{\cdot}002$) previously measured in diluted potassium chromium selenate alum (Bleaney and Bowers 1951). The sign of D is found from the change in the relative intensity of the outer lines with temperature (see Bleaney, Scovil and Trenam 1954); no change in the parameters D, E between 90° and 13°K was observed. Hebb and Purcell (1937) found it necessary to assume that D is negative in potassium chromic sulphate in order to interpret the results of adiabatic demagnetization experiments; Hudson and McLane, and Gardner and Kurti have made the same assumption for the methylamine alum.

Parameter	Diluted Salt	Undiluted Salt
g	$1{\cdot}977 \pm 0{\cdot}003$	$1{\cdot}976 \pm 0{\cdot}007$
D	-958 ± 4	-871 ± 7
E	-92 ± 8	-92 ± 8
$2(D^2+3E^2)^{1/2}$	1942 ± 10	1770 ± 20
δ	1939 ± 6	1790 ± 30

The spin Hamiltonian parameters and the Stark splitting at liquid hydrogen temperatures. The constants D, E and the splitting δ are given in units of 10^{-4} cm^{-1}; $2(D^2+3E^2)^{1/2}$ is the splitting calculated from measurements in strong fields, and δ is the splitting determined directly by extrapolation from small fields to zero field.

3.2. *Measurements at Low Fields*

The specific heat below 1°K depends upon the separation δ of the doublets in zero field, but the value of δ calculated from parameters measured at high fields involves an extrapolation which is only valid if the chosen Hamiltonian is correct. In order to check this, measurements were made at microwave frequencies which were close to the separation δ. The spectrum then has lines very close to zero field, and a plot of the fields in which the lines occur as a function of the frequency enables one to extrapolate to zero field to find an approximate value of δ. When the microwave frequency is set at this value only slight adjustments of frequency are required to bring the lines precisely into zero field; measurement of the frequency then gives δ to high precision. The value of δ measured in this way on the dilute sample gave excellent agreement with that deduced from the high field parameters (see table).

It was not, however, possible to do this with the undiluted salt as the lines were so much broader. This, together with the fact that there are several lines converging on the same frequency whose relative intensities change as they approach zero field, made it extremely difficult to judge precisely when the line was in zero field. The absorption of the microwaves will decrease most rapidly as the field is increased when the frequency is exactly equal to δ, but this cannot be judged precisely because of the changing relative intensity of the component lines contributing to the absorption. Thus it is probably more reliable to extrapolate from small fields where the expected dependence of line position upon frequency can be calculated.

Several measurements have been made: (*a*) magnetic field parallel to $\{111\}$: extrapolating from small fields at frequencies above and below δ (figure 2) gives $0{\cdot}181 \pm 0{\cdot}005$ cm^{-1}; (*b*) field parallel to $\{111\}$: the line in zero field gives $0{\cdot}177 \pm 0{\cdot}010$ cm^{-1}; (*c*) field parallel to $\{100\}$: at frequencies greater than δ

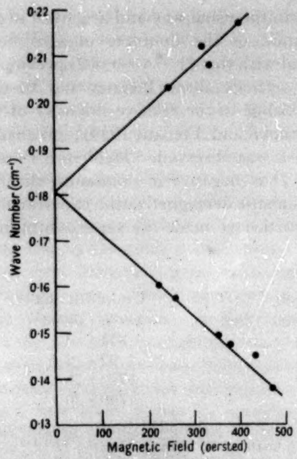


Figure 2.

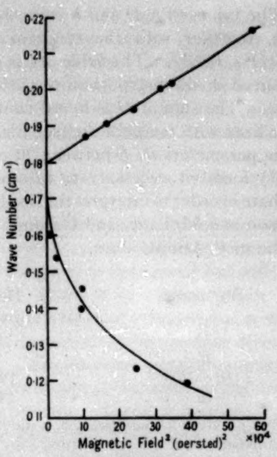


Figure 3.

Figure 2. The wave number at which a line occurs as a function of the magnetic field when the field is parallel to a {111} direction. A linear Zeeman effect is obtained with equal values of $(d\bar{\nu}/dH)$ in the two branches.

Figure 3. The square of the magnetic field at which the line occurs as a function of wave number when the field is parallel to a {100} direction. A simple quadratic Zeeman effect is found for one branch : the other shows a mixture of linear and quadratic effects.

(figure 3) this gives $0{\cdot}180 \pm 0{\cdot}005$ and at frequencies lower than δ (figure 3) this gives $0{\cdot}176 \pm 0{\cdot}010\,\mathrm{cm}^{-1}$. The variation with frequency of the field at which lines were observed was in good accord with the functions calculated from equation (2). The weighted mean of these values of δ is $0{\cdot}179 \pm 0{\cdot}003\,\mathrm{cm}^{-1}$ which agrees well with that calculated from the high field parameters $0{\cdot}177 \pm 0{\cdot}002\,\mathrm{cm}^{-1}$.

It is of interest to note that when $\bar{\nu} = 0{\cdot}187\,\mathrm{cm}^{-1}$, the value of δ required by the results of Hudson and McLane, the lowest line is at a field of over 300 oersteds.

§ 4. Discussion

4.1. *The Nature of the Transition*

The temperature of the transition, which was determined while the temperature of the specimen was rising, was 170 ± 2°K and 157 ± 2°K for the diluted and undiluted crystals respectively. These temperatures correspond very closely to those of 170 ± 2°K and 160 ± 2°K at which Griffiths and Powell (1952) found an abrupt discontinuity in the dielectric constant of aluminium and chromic methylamine alums respectively.

Nuclear resonance of the protons in aluminium methylamine alum has been observed by Rollin and Watson (unpublished). At room temperature and at 195°K there was a sharp line, but at 90°K there was only a broad line (about 30 oersteds). In order that an axially symmetrical ion like methylamine may be accommodated on a lattice site normally occupied, in other alums, by spherically symmetrical ions it must probably undergo a random tumbling motion to produce a time averaged spherical symmetry and, also, a narrow resonance line. At low temperatures its energy of tumbling would not be sufficient to surmount the potential barrier and it would be frozen in to some relatively fixed orientation, giving a broad line. This process might necessitate a lowering of the symmetry of the structure to accommodate the axially symmetrical radical. All of the available results are consistent with such an explanation, but it would be of interest to carry out X-ray structure measurements at low temperatures to verify that this is the correct explanation. The presence of this transition may explain the effects obtained by Hudson and McLane which depend upon the rate at which the crystal is cooled to liquid nitrogen temperatures.

4.2. *The Discrepancy between Measurements of* δ

This discrepancy is conveniently discussed in terms of the specific heat C of the spin system, which at high temperatures may be written in the form

$$CT^2/R = \delta^2/4 + 2{\cdot}4\tau^2 + \Delta^2/8 \qquad \ldots\ldots(3)$$

where the second and third terms represent contributions from magnetic and isotropic exchange interaction respectively. Δ is the Curie–Weiss constant, which Gardner and Kurti report as $0 \pm 0{\cdot}02$°K, so that this contribution is very small (10^{-4} deg^2) and can be neglected. The value of τ is $0{\cdot}02$°K and hence magnetic interaction contributes 10 (in units of 10^{-4} deg^2) while the paramagnetic resonance value of δ gives $\delta^2/4 = 164$, so that the expected total is 174. The thermal measurements, on the other hand, give a total of 193, from which δ was deduced by subtracting the magnetic contribution from the measured total (the error in this total is ±14 (Gardner and Kurti) and ±4 (Hudson and McLane), as computed from the errors given for δ). Hence there is excess specific heat amounting to 19 in these units, which might be due to the presence of anisotropic exchange interaction. The second term in equation (3) would then have to be replaced by a more general expression, but Δ would remain unaltered (see Baker and Bleaney 1955). Such interaction would contribute to the mean square width of the paramagnetic resonance spectral lines, and some measurements of this quantity were attempted.

When the external magnetic field is parallel to the [111] direction the line in highest field (the $+3/2 \longleftrightarrow +\frac{1}{2}$ transition for the ion whose axis is parallel to the field) is well clear of the rest of the spectrum so that a reliable line shape measurement can be made on it. The root mean square width, measured at both 90°K and 20°K, is 116 ± 6 oersteds compared with 116 oersteds calculated from the magnetic interaction with neighbouring chromic ions, assuming that they are in the same positions that they occupy at room temperature. Also the ratio of the measured half width at the half power points to the root mean square width is $1{\cdot}26 \pm 0{\cdot}07$ compared with 1·18, the value of this ratio for a gaussian line shape. Van Vleck (1948) shows that for a simple cubic lattice with the magnetic field parallel to a body diagonal the line shape should be nearly gaussian for purely magnetic interaction, and Bleaney (1950) verified this for the corresponding line in caesium chrome alum.

The neighbours of any chromic ion can be regarded as dissimilar as they have different magnetic axes. Van Vleck has shown that for this case isotropic exchange will contribute to the root mean square width of the line. Also anisotropic exchange will in general contribute to the root mean square width. Thus these measurements give no indication of exchange, but the possibility remains that exchange effects produce a change in line width in other directions. This could not be checked in the concentrated salt as line width measurements in other directions are impossible because the spectral lines overlap one another. However, useful measurements can be made in a semidilute crystal of aluminium methylamine alum containing 5 to 10% chromium. Such a crystal displays, in addition to the usual spectrum due to single ions, a weaker spectrum due to pairs of interacting ions which is centred on each of the spectral lines due to isolated ions, on either side of which they appear as satellites; the separation of the satellites is proportional to the interaction between the neighbours. Thus the measurement of the satellite separation in any direction gives the strength of the interaction tensor in that particular direction.

When the field is parallel to the [111] direction the extreme satellites are visible on the central line of the spectrum, and they are separated by 45 ± 5 oersteds compared with 40 oersteds expected from the magnetic interaction alone. This is in agreement with the fact that the mean square width in the undiluted salt is that expected from magnetic interaction alone. When the field is parallel to the [110] direction the extreme satellites are visible on the two central lines of the spectrum (the central line for two of the ions is separated from that for the other two by second order effects), and are separated by 150 ± 10 oersteds compared with 158 oersteds expected from magnetic interaction alone. In other directions the satellites are not visible because they are masked by the much larger lines due to isolated ions. It is unlikely that the interaction tensor would be equal to the magnetic interaction in two directions and yet be different in the third direction, although this is of course a possibility. It therefore seems probable that there is only very small exchange interaction between the chromic ions and that the discrepancy between the paramagnetic resonance and other measurements cannot be explained in this way.

There remains the possibility that the energy levels of the methylamine ion have a small splitting, such as might be associated with resonance between several equivalent positions. To explain the excess specific heat would require a splitting of the order 0·08°K, with which would be associated a Schottky anomaly at a few hundredths of a degree Kelvin. No trace of such an anomaly appears in the measurements of Gardner and Kurti, Hudson and McLane or Beun, Steenland, de Klerk and Gorter. Also, the close agreement between the relaxation measurements of Cooke and Wolf and the calorimetric measurements shows that the excess specific heat resides in the spin system rather than the lattice, whereas the methylamine ions are much more likely to be coupled to the latter than to the former.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to the Warden and Fellows of St. Antony's College, Oxford for a Senior Scholarship while this work was being done. He is considerably indebted to Dr. B. Bleaney and to Dr. N. Kurti for helpful discussion of results.

REFERENCES

BAGGULEY, D. M. S., BLEANEY, B., GRIFFITHS, J. H. E., PENROSE, R. P., and PLUMPTON, B. I., 1948, *Proc. Phys. Soc.* A, **61,** 542.

BAGGULEY, D. M. S., and GRIFFITHS, J. H. E., 1950, *Proc. Roy. Soc.* A, **204,** 188.

BAKER, J. M., and BLEANEY, B., 1955, *International Low Temperature Conference, Paris.* (Paris : Institut International du Froid).

BEUN, J. A., STEENLAND, M. J., DE KLERK, D., and GORTER, C. J., 1955, *Physica,* **21,** 767.

BLEANEY, B., 1950, *Proc. Roy. Soc.* A, **204,** 203.

BLEANEY, B., and BOWERS, K. D., 1951, *Proc. Phys. Soc.* A, **64,** 1135.

BLEANEY, B., SCOVIL, H. E. D., and TRENAM, R. S., 1954, *Proc. Roy. Soc.* A, **223,** 18.

GARDNER, W. E., and KURTI, N., 1954, *Proc. Roy. Soc.* A, **223,** 542.

GRIFFITHS, J. H. E., and POWELL, J. W., 1952, *Proc. Phys. Soc.* A, **65,** 289.

HEBB, M. H., and PURCELL, E. M., 1937, *J. Chem. Phys.*, **5,** 338.

HUDSON, R. P., and MCLANE, C. K., 1954, *Phys. Rev.*, **95,** 932.

LIPSON, H., 1935, *Proc. Roy. Soc.* A, **151,** 347.

LIPSON, H., and BEEVERS, C. A., 1935, *Proc. Roy. Soc.* A, **148,** 664.

VAN VLECK, J. H., 1948, *Phys. Rev.*, **74,** 1168.

Variation of Contrast in Electron Microscope Images with size of Objective Aperture

By J. H. COUPLAND
Wheatstone Physics Laboratory, King's College, London

Communicated by J. T. Randall; MS. received 17*th October* 1955 *and in revised form* 22*nd December* 1955

Abstract. Measurements of the contrast in electron microscope images show a much greater variation with size of objective aperture than predicted by some of the earlier calculations. Experimental results for an inorganic object are compared with values derived from current scattering theory for the angular range $10^{-3}<\alpha<10^{-2}$ rad, where α is the angle of scattering. Better agreement is obtained, but a significant discrepancy remains.

§ 1. Introduction

It is well known in practice that the contrast in electron microscope images changes considerably with variation of relative objective aperture over the range $10^{-3}<\alpha<10^{-2}$ rad. If axially parallel illumination is assumed then α represents the maximum angle of deflection of electrons scattered at the specimen and accepted by the objective to contribute to the final image. Estimates of atomic cross sections for scattering at very small angles have been made by Marton and Schiff (1941), Boersch (1947), von Borries (1949) and Hillier and Ramberg (1949), but none appears to be in full agreement with experiment. Biberman and associates (1949) have drawn attention to the rapid increase in cross section at small angles and conclude that it must be due to inelastic processes, though in this particular case Lenz (1954) has been able to modify the elastic scattering theory to account for their experimental values for chromium. The work of Leonhard (1954) further supports the Biberman suggestion.

By taking a multiply folded 'Formvar' film as an organic test object, contrast measurements were made at 150 kv for several angles α and different thicknesses, and the results were mentioned very briefly at the Electron Microscopy Conference, London 1954 (Coupland 1954, 1955). Figure 1 illustrates the effect. Since these earlier measurements did not conform to theoretical predictions, see figure 2, they have been repeated with greater accuracy and under slightly different conditions to give substantially the same results. An account of this work now follows.

§ 2. Experimental Method

'Formvar' films were prepared on glass slides from an ethylene dichloride solution by the method of Revell and Agar (1955) so as to obtain a near uniform film thickness. Precautions were taken to minimize film shrinkage. No significant change in thickness was found on using a divided beam interference microscope on the specimen before and after examination in the electron microscope, but the accuracy of the method as used was limited to ± 10%. The film thickness was measured by the multiple beam technique in reflection as developed by Tolansky (1948), and by combining this with the measured density ($1{\cdot}24\,\mathrm{g\,cm^{-3}}$) of 'Formvar' cast from solution the mass per unit area of film was obtained.

2.1. *Photographic Technique*

The ratio of the incident electron beam intensity I_0 to that of the transmitted beam I_c was obtained by measuring the blackening on appropriate areas of a suitable micrograph of known calibration. By using recording film instead of plates thirty exposures may be made on a single film, thus permitting a time-graded series at constant intensity to be made on the same film for three different values of α. A characteristic (H,D) plot of optical density D against $\log_{10}(It)$, where (It) represents the exposure, gave a linear portion with slope $Y=3$ for Kodak Microfile film. This value of Y will be inappropriate to the analysis if reciprocity in I and t is not obeyed (cf. Digby *et al.* 1953). To check this point, a similar exposure sequence was taken for different known intensities obtained by variation of the current in the intermediate projector lens controlling the magnification. Identical (H,D) plots were obtained for both series over the linear portion and hence reciprocity may be assumed, but some differences were observed at low densities.

2.2. *Variation of* α

In the high voltage electron microscope (Coupland 1954, 1955) used for these measurements, two standard type copper aperture discs are mounted on a copper rod which may be moved through the non-magnetic spacer in the pole-piece gap of the objective lens. This allows for two alternative discs or none to be employed without breaking the vacuum system. The values of α are then taken as $\alpha=[(\text{diameter of aperture})/2f_0]+e$, with the correction e ($=+7\%$) to allow for the curvature of the electron trajectory and the extra focal position of the object. The focal length f_0 ($=0{\cdot}74$ cm) was deduced from the magnification of the first stage. Care was taken to ensure that the maximum semi-angle of convergence of the illumination did not exceed 10^{-4} radian.

§ 3. ANALYSIS OF RESULTS

All measurements of photographic density were made over extended areas at low magnification so that lens errors and edge contour effects were unimportant. Only in the absence of a limiting objective aperture, i.e. very large α, was there any general background scattering over the whole field of view as measured by the intensity in an image region corresponding to an opaque object. Hall (1951) found that it was possible, by using a dark field mask placed a few millimetres below the specimen, to separate the transmitted electron beam I_c, which is accepted by the objective aperture, into two components, I_t the residual and effectively undeflected beam, and I_k the portion deflected through angles less than α, so that $I_c=I_k+I_t$. In the experiments with a film 500 Å thick, where the probability of scattering is small for 150 kv electrons, it was found difficult to distinguish I_t from I_k over the angular range $10^{-3}<\alpha<10^{-2}$ radian.

Figure 1 shows within experimental error that the usual exponential relation $I_c=I_0\exp(-n\sigma'x)$ is applicable. Here x is the film thickness, σ' is the cross section for single scattering through an angle greater than α and n is the number of atoms/cm³. It has also been shown by multiple scattering analysis

(Zworykin *et al.* 1945) that this relation would still hold experimentally if the film thickness and angles are sufficiently small as is the case. However, on the basis of this equation, values of σ' may be compared and plotted as a function of α.

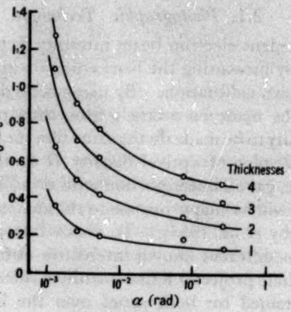


Figure 1. Logarithmic plot of contrast with aperture angle for folded ' Formvar ' film.

§ 4. COMPARISON WITH THEORY

4.1. *Elastic Scattering*

According to the theory of Molière (1947), the elastic scattering per unit solid angle Ω is given by

$$\frac{d}{d\Omega}\left(\frac{N}{N_0}\right)=40{\cdot}8\left(1+\frac{u}{511\,000}\right)^2\frac{Z^{2/3}}{A}\left(\sum_{i=1}^{3}\frac{a_i}{b_i^2+(\theta/\theta_0)^2}\right)^2(\gamma x)$$

$$\theta_0=4{\cdot}17\,Z^{1/3}u^{-1/2}\left(1+\frac{u}{1\,022\,000}\right)^{-1/2}\quad u=150\,000\text{ volts,}$$

$$\theta=\text{angle of scatter, }(\gamma x)=\text{'mass thickness' }(\mu g\text{ cm}^{-2})$$

and

$i=$	1	2	3
$a_i=$	0·1	0·55	0·35
$b_i^2=$	36	1·44	0·09

Using these formulae, values of $(d/d\theta)(N/N_0)$ for carbon were obtained and subsequent integration then gave the elastic cross section σ_e for scattering through angles greater than α. In figure 2, σ_e for 10^{-6} g of material is plotted as a function of α.

If different elements (including the special case of hydrogen) be considered then on most theories of scattering σ_e, expressed per unit mass, varies only slowly with atomic number. As experimental evidence the work of Hall (1955) on a wide range of elements may be cited, but in this case the appropriate angle α was large and uncertain as no limiting aperture appeared to be used. Von Borries (1949) has shown that the scattering formulae of Debye, Bethe and Molière all give substantially similar results for σ_e, and it may be shown that these agree equally well with values derived from data by Bullard and Massey (1930) for a Thomas–Fermi atom when a full relativity correction factor for mass and wavelength is applied.

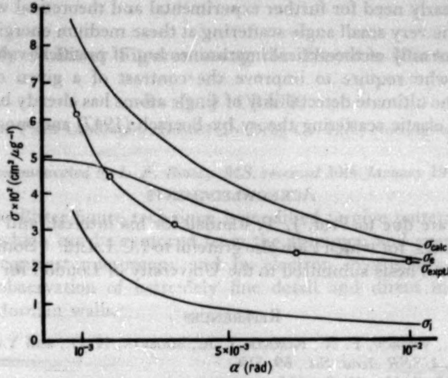


Figure 2. Plot of σ_i, after Koppe, for carbon, σ_e, after Molière, for carbon, $\sigma_{calc}=\sigma_i+\sigma_e$, σ_{exp} as by experiment, for 'mass thickness' 1μ gcm^{-2} ($u=150\,000$ v).

4.2. *Inelastic Scattering*

Values for the inelastic cross section σ_i for carbon have been evaluated from the theory of Koppe (1948) and plotted in figure 2 together with total values, $\sigma_{calc}=\sigma_e+\sigma_i$. The simpler formulae of Marton and Schiff (1941) show a similar variation for σ_i over this angular range. According to Koppe

$$\frac{d}{d\Omega}\left(\frac{N}{N_0}\right)=0{\cdot}830\;\frac{1}{AZ^{5/3}}\left(1+\frac{u}{511\,000}\right)^2\frac{G(V)}{V^3(V+V^*)}\quad(\gamma x)$$

where $$V=0{\cdot}0902\;\frac{u^{1/2}}{Z^{2/3}}\left(1+\frac{u}{1\,022\,000}\right)^{1/2}\left[\left(\frac{I}{4u(1+u/1\,022\,000)}\right)^2+\theta^2\right]^{1/2}$$

and $$V^*=\frac{1{\cdot}68\times10^{-6}}{|\chi|Z^{1/3}}.$$

The function $G(V)$ of Bewilogua–Heisenberg is tabulated by von Borries (1949, p. 64), I is the ionization energy (volts) and χ is the diamagnetic atomic susceptibility (cm³).

4.3. *Total Scattering*

The average experimental values σ_{exptl} for the total scattering are also plotted in figure 2 for comparison with σ_{calc} which assumes that the film is amorphous. There is only a rough order of magnitude agreement, and the differences in curvature imply corresponding discrepancies in total differential cross section. In the experimental results the angular variation is accurate to ± 5%, but there may be a systematic error in absolute values of the order of 10% due to shrinkage or irregularities in film thickness.

The recent extensive work of Leonhard (1954) on thin films of beryllium, aluminium and gold is very relevant. By using a velocity spectrometer and 50 kv electrons he has separated the components and finds that the experimental elastic scattering is far less than the Molière theory suggests. The measured inelastic scattering, however, appears to follow the theory of Koppe for beryllium, but is much greater than this theory predicts for aluminium and gold.

There is clearly need for further experimental and theoretical work for a full evaluation of the very small angle scattering at these medium energies. Such an evaluation is not only of theoretical importance but of practical value to electron microscopists who require to improve the contrast of a given object. As a limiting case, the ultimate detectability of single atoms has already been discussed on the basis of elastic scattering theory by Boersch (1947) and more recently by Haine (1954).

ACKNOWLEDGMENTS

My thanks are due to Prof. J. T. Randall for his interest, and for providing laboratory facilities, for which I am also grateful to I.C.I. Ltd. Some of this work formed part of a Thesis submitted to the University of London for the degree of Ph.D.

REFERENCES

BIBERMAN, L. M., VTOROV, F. N., KOVNER, I. A., SOUSKIN, N. G., and YAVORSKI, B. M., 1949, *Rep. USSR Acad. Sci.*, **69,** 519.
BOERSCH, H., 1947, *Z. Naturf.*, **2a,** 615.
VON BORRIES, B., 1949, *Z. Naturf.*, **4a,** 51.
BULLARD, E. C., and MASSEY, H. S. W., 1930 *see* MOTT, N. F., and MASSEY, H. S. W., *Theory of Atomic Collisions* (Oxford : University Press), p. 190.
COUPLAND, J. H., 1954, *Ph.D. Thesis*, University of London; 1955, *Proceedings of the International Conference on Electron Microscopy, London*, 1954 (London : Royal Microscopical Society).
DIGBY, N., FIRTH, K., and HERCOCK, R. J., 1953, *J. Photogr. Sci.*, **1,** 194.
HAINE, M. E., 1954, *Advances in Electronics*, **6** (New York : Academic Press).
HALL, C. E., 1951, *J. Appl. Phys.*, **22,** 655 ; 1955, *J. Biophys. and Biochem. Cytology*, **1,** 1.
HILLIER, J., and RAMBERG, E. G., 1949, *Proceedings of the Electron Microscopy Conference, Delft* (The Hague : Nijhoff), p. 42.
KOPPE, H., 1948, *Z. Phys.*, **124,** 658, *see* von Borries, 1949.
LENZ, F., 1954, *Z. Naturf.*, **9a,** 185.
LEONHARD, F., 1954, *Z. Naturf.*, **9a,** 727,1019.
MARTON, L., and SCHIFF, L. I., 1941, *J. Appl. Phys.*, **12,** 761.
MOLIÈRE, G., 1947, *Z. Naturf.*, **2a,** 133.
REVELL, R. S. M., and AGAR, A. W., 1955, *Brit. J. App[illegible]ys.*, **6,** 23.
TOLANSKY, S., 1948, *Multiple Beam Interferometr[illegible] Surfaces and Films* (Oxford : University Press).
ZWORYKIN, V. K., *et al.*, 1945, *E[illegible]ptics and the Electron Microscope* (New York : John Wiley).

A Study of Bitter Figures using the Electron Microscope

By D. J. CRAIK

The Plant Laboratory, Boots Pure Drug Co. Ltd., Nottingham

Communicated by L. F. Bates; MS. received 30th January 1956

Abstract. The Bitter figure technique is modified to give patterns in the form of thin films which can be removed from the specimen and examined by high power phase-contrast microscopy and by electron microscopy. This makes possible the observation of extremely fine detail and direct measurement of the width of domain walls.

---

§ 1. INTRODUCTION

THE well-known powder pattern method for the study of domain configurations in ferromagnetic materials (Bitter 1931) consists in placing a drop of a colloidal suspension of magnetite on the specially prepared surface of the ferromagnetic. Where domain boundaries (Bloch walls) intersect the surface, strong local magnetic fields exist and attract the magnetic particles of the colloid so that in the majority of cases the position of the domain walls can be observed.

It is not considered that the pattern formed is a true representation of the wall itself. Commonly the amount of colloid used is such as to give deposits of width around 10^{-4} to 10^{-3} cm, which is several times the width ascribed to the wall. If a less concentrated colloid is used the deposits are correspondingly finer, but a limit is soon reached where the pattern can no longer be clearly observed by reflection microscopy.

The object of this investigation was to modify the method in such a way that much finer patterns could be clearly observed, and ultimately to prepare patterns which could be considered to be more direct representations of the domain walls. These latter patterns had to be prepared in such a way that they could be observed in the electron microscope, facilitating measurements down to 10^{-6} cm.

§ 2. METHOD

The new method consists in the preparation of thin films of the patterns which can be dried without distortion and may be removed from the metal surface for examination by high power transmission microscopy, or by electron microscopy. In a stable colloid the particles are held apart by electrostatic repulsion, arising from the inherent charges of the particles themselves or from layers of adsorbed ions. When the volume of the disperse phase is reduced the particles are forced into close proximity and, when drying takes place, they adhere to form coarse aggregates. The appearance of a normal magnetite suspension, after drying, is shown in figure 1.†

† For figures see Plates at end of issue.

It was clearly necessary to produce a colloid which would show an entirely different behaviour when dried, and it was therefore necessary to incorporate in the colloid a substance which would satisfy the following conditions. First, it must act as a protective colloid to maintain the particles in a highly dispersed state. Secondly it must, on drying out from a weak solution, show a gradual increase in viscosity until it eventually solidifies without any abrupt change of state and locks the particles in position, i.e. it must be amorphous. And finally it must form a film which is regular and tough enough to withstand manipulation and the heating effects of the electron beam, although it is only a few hundred ångströms thick.

The substance eventually chosen was a methyl ether of cellulose known by the trade name ' Celacol '. The magnetite was prepared in a finely particulate but aggregated form by the action of caustic soda on ferrous and ferric chlorides. The deposit was filtered and washed with distilled water. Hydrochloric acid was added to peptize the deposit sufficiently for it to be carried through a fine filter paper. The ' Celacol ' was added at this stage to 0·1%. In order to remove the acid the colloid was dialysed against several changes of distilled water until no chloride reaction was given. If the acid was removed before the protective colloid was added complete flocculation took place—the magnetite by itself did not form a stable suspension. The final preparation was quite stable and could be used for several weeks, although it was considered advisable to re-filter it before use.

When a few drops of this colloid were spread on a polished stainless steel surface and allowed to dry, a thin tough film was formed. When this film was stripped and mounted on a glass slide, it was found on microscopic examination to be quite featureless with no visible particles on aggregates. Electron microscopic examination of such a film showed particles of a few hundred ångströms diameter uniformly distributed in the film (figure 2). If a structure is found in such a film it is assumed to be evidence of the existence of magnetic fields which vary from place to place. According to Kittel (1949) the field necessary to form a pattern when the particle diameter is 500 ångströms is 2 oersteds, while smaller particles need some 300 oersteds. The general method of preserving the state of dispersion of aqueous colloids when dried appears to solve an outstanding problem in electron microscopy and will be published in greater detail elsewhere.

In the investigation of cobalt specimens no problems of corrosion arise. If the colloid dries directly on the surface, however, the resulting film is hard to remove intact, and accordingly the following procedure was used. A film of Formvar resin some 200 ångströms thick was formed by dipping a glass slide into a 0·1% solution of Formvar in ethylene dichloride and allowing it to dry in a vertical position. The slide was then placed in a petri dish containing distilled water in such a way that the film floated off at one end but remained attached at the other. By means of angled forceps the cobalt specimen was placed below the film and on the slide, which was then withdrawn so that the film settled down evenly on the specimen.

To form the pattern, a measured quantity of the colloid was placed on the film-covered surface and spread out with a bent platinum wire, care being taken not to scratch the film. When the pattern was intended for electron microscopy the quantity used was sufficient to give a film some 500 ångströms thick, while the strength of the magnetite was varied to give deposits of varying density. Usually the colloid was almost colourless, with a faint amber tint. To prepare a pattern for light microscopy, a stronger colloid was used to give both a heavier deposit and a thicker film.

After the colloid had dried, a few drops of a thick solution of collodion in amyl acetate were applied. This dried to form a relatively thick film which was peeled off, carrying the pattern-bearing film with it, by means of ' Sellotape ', or by using forceps. With reasonable care the pattern was not distorted during this process. For optical observation the patterns were mounted under a cover-slip with Canada balsam, when they formed a permanent record and could be examined and photographed at will. For electron microscopy, selected areas of the composite film were laid on the standard copper grids with the collodion side downwards. The grids rested on curved strips of stainless steel mesh, below which a quantity of amyl acetate could be introduced in such a way that it came into contact with the collodion. This amyl acetate was withdrawn and renewed at intervals until the collodion was completely dissolved and the thin film of ' Celacol ' remained stretched across the grid. The procedure resulted in very little distortion or folding of the film.

When this method was applied to iron specimens a major consideration was the rusting which would have ensued if the colloid dried directly on the surface. This was circumvented by covering the surface with a very thin film of quartz, which was deposited by vacuum evaporation. A film even a few hundred ångströms thick gave complete protection. Since some deterioration of the surface takes place when the liquid colloid is used, it is suggested that this method of protection would be of value when the conventional methods of observation are used.

§ 3. EXPERIMENTAL RESULTS

Patterns on two cobalt crystals have been investigated. Micrographs of a pattern prepared for optical microscopy, using a cobalt disc, are shown in figures 3 and 4, where 180° walls are shown, and the very intricate pattern on a face perpendicular to the hexagonal axis is represented in considerable detail. In order to show the fainter lines without coarsening the general pattern by the application of an excessive amount of colloid, phase-contrast microscopy was used. This was found to be of very great advantage.

Patterns were prepared for electron microscopy with two objects: to elucidate the very fine detail of certain patterns, and to obtain information about the domain walls themselves. A cobalt strip with a face cut perpendicular to the hexagonal axis yielded an extensive pattern of which a detail is shown in figure 5, and at a higher magnification in figure 6 ; the advantage of the very high contrast, inherent in electron microscopy, becomes apparent. Figure 7 shows the distribution of particles across part of a dagger-shaped domain on a cobalt crystal whose surface made a small angle with the hexagonal axis. Figure 8 is a typical detail from a lace pattern formed on the (100) surface of a silicon–iron crystal with an applied field of 40 oersteds in the direction shown. The gross pattern was made up of long, roughly parallel lines of such features, perpendicular to the direction of the applied field.

Examples of the finest patterns obtained are shown in figures 9–13. Figures 9 and 10 show 180° walls on a silicon–iron (100) face. A very weak colloid was used. Figure 9 includes a striation, indicating that the magnetization within the domain is parallel to the wall. Although the deposit is somewhat irregular, rough measurements gave the average width of such lines to be between 2 and 5×10^{-5} cm, agreeing quite well with the theory.

Figures 11, 12 and 13 are the results of a systematic attempt to determine the width of the domain wall on a face parallel to the hexagonal axis of cobalt. Successively more dilute colloids were used, the concentration of the 'Celacol' and the quantity used being kept constant. In figure 11 the width of the deposit is approximately 4×10^{-5} cm. Figure 12 shows part of one of several lines of which the average width was $1{\cdot}5 \times 10^{-5}$ cm. Finally, figure 13 is an example of a pattern obtained with an almost colourless colloid, the average width being 8×10^{-6} cm. In figure 11 the line is seen to be very regular and unbroken, as is to a lesser extent the line of figure 12. When the amount of magnetite is reduced to the extent represented by figure 13, small gaps begin to appear and the line is less regular. When even more dilute colloids are used, isolated particles and clumps appear spread out along the lines.

Clearly, reducing the amount of magnetite available reduces the width of the deposit to an extent where there is insufficient to cover the whole of the domain wall. It is suggested that when this stage is reached the width of the deposit corresponds with the width of the wall itself, the value in this case being approximately 1×10^{-5} cm.

At present it is difficult to draw any conclusions as to the configuration of the field within the walls. In order to gain such information an even finer colloid would be desirable, and certain modifications might be made to the technique. It might then be found that a photometer trace across a very high magnification electron micrograph would lead to the experimental determination of the field distribution. However, examination of the figures does suggest that the distribution is rather flat with a fairly sharp decrease at the edges.

ACKNOWLEDGMENTS

I most gratefully acknowledge the advice and encouragement of Professor L. F. Bates of the University of Nottingham, and of Mr. W. J. Randall, Manager of the Plant Laboratory. My thanks are also due to Mr. D. M. Riekie, for his advice on colloids, to Mr. D. A. Meads and Mr. B. F. Miller, for their technical assistance, and to Dr. A. Hart and Mr. P. F. Davies, for polishing the specimens. The cobalt disc referred to in § 3 was supplied by Professor W. Sucksmith.

REFERENCES

BITTER, F., 1931, *Phys. Rev.*, **38**, 1903.
KITTEL, C., 1949, *Rev. Mod. Phys.*, **21**, 541.

Pre-Breakdown Current and Vacuum Breakdown

By the late W. J. R. CALVERT

Queen Mary College, University of London

MS. received 6th January 1955, and in revised form 30th December 1955

Abstract. An account is given of measurements of the development of current between un-outgassed electrodes at distances up to 4·5 cm with uniform electric fields at voltages up to 100 kv at pressures of the order of 10^{-5} mm Hg. Four distinct stages of development were noted : (i) an initial pre-breakdown stage in which a cold emission of electrons under electric fields of the order of 10^5 v cm^{-1} took place; (ii) the occurrence of current pulses of the order of 10^{-7}c at a critical voltage; (iii) larger, repetitive pulses of the order of 10^{-5}c at a slightly higher voltage accompanied by x-ray generation and the release of gas, and (iv) complete breakdown in which the current was determined by circuit constants at a still higher voltage. The pulse threshold voltage for large gaps was determined by the total gap voltage. The development of the field emission initiated arc is discussed in terms of the particle exchange theory and the importance of x-rays and photons and of the release of gas are discussed.

§ 1. Introduction

The electrical breakdown of vacuum insulation is a problem of considerable practical importance owing to the increasingly widespread application of so-called vacuum insulation in high voltage devices such as particle accelerators, x-ray tubes, high frequency low-loss condensers, etc. Owing to their large physical size it is often impracticable effectively to outgas such apparatus, which may therefore operate at pressures of residual gas of the order of 10^{-5} mm Hg. The present work is a preliminary investigation of what is called vacuum breakdown under such conditions, and consists of an examination of the development of current between electrodes at pressures of the order of 10^{-5} mm Hg and with electric fields of the order of 10^5 v cm^{-1}. The general problem is to account for the initiation and broad action of a conducting electrical discharge in these conditions.

The build-up of current leading to vacuum arcs has been the subject of many previous investigations (see for example Langmuir 1913, Coolidge 1913, Bennet 1932 a, b, 1933, 1934, Ahearn 1936, Cranberg 1952 and Kilpatrick 1953), and this previous work has suggested that field emission may be responsible for pre-breakdown currents at potentials greater than about 10 kv and it is generally recognized that field emission can play an important part in the initiation of breakdown. The recent comprehensive work of Dyke *et al.* (1953b) and Dolan *et al.* (1953), using clean outgassed electrodes under very high vacuum conditions, has established that as the field emission current density is increased this emission is terminated by an explosive arc, the arc being initiated at a critical value of the field emission current density of the order of 10^8 A cm^{-2} and involves both high fields and high temperatures at the cathode. The transfer of material from the anode would appear to play no part in the process.

The conditions on which higher residual gas pressures of the order of 10^{-5} mm Hg prevail and with un-outgassed electrodes used in practical applications are often more complex, and it is important to find what mechanism or mechanisms operate in these conditions. The precise mechanism of the transition from low current conduction to a high current arc in these circumstances is not fully understood, but a number of mechanisms involving the transfer of anode material, impact of ions and photons, the action of adsorbed gases and residual gas have been suggested to account for breakdown.

Moreover, it is not always clear from earlier work that a number of different types of discharge may occur. This fact has led to some confusion and may account for apparent discrepancies between the work of different observers. It is desirable, therefore, to describe the phenomena which were observed in the present investigation. The experimental technique adopted involved the measurement of currents flowing across the so-called vacuum gap, and their dependence on gap voltage and nature of the electrode surfaces.

§ 2. Apparatus

The apparatus used consisted of two plane parallel copper electrodes A and C mounted in an evacuated discharge chamber D of internal diameter 10 cm (see figure 1) and capable of separation up to distances of about 4 cm, a source of

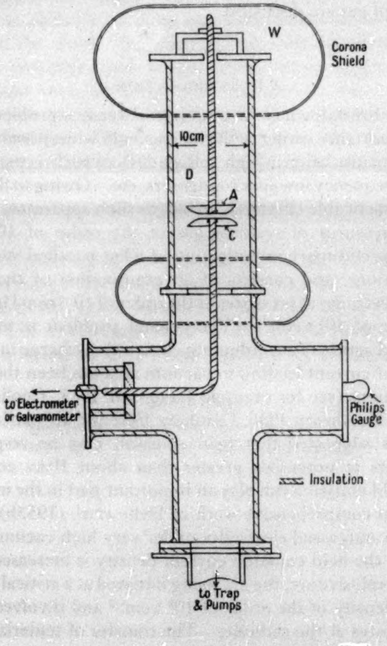


Figure 1. Discharge chamber.

stabilized high voltage, and current and voltage measuring apparatus. A Geiger counter connected to amplifiers and a loudspeaker was mounted alongside the chamber in order to detect any X-ray emission.

The discharge chamber used in preliminary measurements was constructed from hard glass and steel, had a volume of 7·5 litres and incorporated a Wilson vacuum seal W which permitted the electrode separation to be altered without disturbing the vacuum. In later work, an improved and larger chamber was used. The electrodes consisted of copper discs 6 cm in diameter with rounded edges to avoid local field distortion and to ensure a uniform field at the centre. Before assembly in the chamber the electrodes were carefully polished with successively finer grades of emery paper, washed and dried, but not given any further treatment. The chambers were evacuated by a mercury diffusion pump and backing pump, and the usual liquid oxygen trap was incorporated to prevent contamination. Pressures were indicated by a Philips gauge which had been calibrated against an ionization gauge. The applied high voltages were measured with a high resistance chain and a sensitive galvanometer, and the usual anti-corona precautions were taken. Inter-electrode currents were measured with galvanometers suitable for currents greater than about 10^{-8} A or (for lower currents) electrometers; the electrometer circuit was screened. In the event of excessive current flowing a neon tube fired and short-circuited the instruments thus protecting them. Voltages up to 100 kV were available.

§ 3. PRELIMINARY EXPERIMENTAL RESULTS

The first experiments were carried out using electrode separation d of 0·5 cm to 2·0 cm to establish the general character of the breakdown phenomena in the parallel plate gap at a residual gas pressure of 2×10^{-5} mm Hg. The broad picture of the train of events is as follows.

Using the galvanometer, no measurable current was found at any given separation as the anode voltage was gradually raised until, at a critical value V_p, a sudden kick was noticed indicating a sudden pulse of current. If the voltage was maintained at this value, the pulse almost immediately extinguished and the magnitude of the charge passed in the pulse was about 10^{-7} C. Reduction of the anode voltage below the initial value eliminated the pulses. The onset of a pulse seemed to depend mainly upon the gap voltage but (to a lesser extent) also on the average field $E(=V/d)$.

Further, there appeared to be at least two types of pulses; one representing discharges of from 1 to 30×10^{-7} C, and the other, much larger and requiring a higher voltage V_p', representing discharges of 10 μC or more. The luminosity produced was also characteristic of the two types. When V_p' was maintained at this higher value a succession of such larger pulses occurred, the repetition frequency and magnitude of the pulses increasing with V_p'. An X-ray generation was recorded by the nearby Geiger counter and this always occurred at the same time as the current pulses in the discharge chamber.

Accompanying the current pulses fluorescence of the glass walls of the chamber of the type associated with electron bombardment was observed. A significant feature with the larger pulses was that when they occurred the Philips gauge gave a sudden large momentary deflection, as if there had been a momentary increase in gas pressure. The pressure might have been appreciable at the cathode, since, despite the response time lag in the pressure gauge and the time thus allowed for the dispersal of the gas, significant increase was actually indicated by the gauge. A typical change of pressure was from about 2×10^{-5} mm Hg to about 8×10^{-5} mm Hg; as the volume of the chamber was about 7·5 litres this would represent the evolution of about 2×10^{16} molecules of gas.

At still larger values of anode potential an apparently continuous, but not necessarily steady, current of several microamperes was observed, the ultimate value being determined by the circuit constants. This continuous current, once started, could be maintained without much reduction in magnitude even though the applied voltage was considerably reduced; the extinction voltage being much lower than the starting voltage. In this, and certain other respects, the discharge resembled a low pressure gas discharge.

A significant result was, however, that with small gaps of the order of 0·1 to 0·2 cm this continuous current stage was reached apparently without the intermediate stages of current pulses. If permitted, the current could increase catastrophically, as in fact it did on one occasion. The continuous currents, although they might be unsteady, were quite distinct from the current pulses. Careful visual observations showed that the apparently continuous discharge consisted in fact of a rapid succession of narrow sparks occurring randomly over the electrode and not confined to any one region. In larger gaps, however, the discharge, if visible at all, was much more diffuse and was largely masked by a marked fluorescence of the glass walls of the chamber. These continuous currents were also accompanied by continuous X-ray generation which was indicated by the Geiger counter, and then appeared to represent the breakdown of the vacuum insulation in the fullest sense.

The fact that no pre-pulse currents could be recorded on the galvanometer, in spite of the apparent influence of the electric field V/d upon the onset of the pulses made it necessary to investigate pre-pulse phenomena in more detail with more sensitive instruments. Accordingly, electrometers were used and these at once disclosed currents of the order of 10^{-11} A to 10^{-9} A, depending on the value of the field E.

§ 4. THE PRE-PULSE CURRENTS

In order to investigate the nature of the origin of the small pre-pulse currents, a large number of measurements was made on the variation of the currents I with gap voltage V for various different gap widths d between 0·15 cm and 2·16 cm. During this stage the well-known conditioning effects (Trump and Van de Graaff 1947) were noted, and difficulty was experienced in obtaining consistently reproducible values of the pre-breakdown currents. This was largely due to the variable conditions to be found with un-degassed electrodes.

Examples of the results for different gap widths are given in figure 2 in the form of graphs of $(\log I/E^2, 1/E)$. The lines 1, 2 and 3 were obtained with a 0·52 cm gap, and line 4 with a 1·03 cm gap. On these graphs the occurrence of a pulse is represented by the sharp vertical peaks.

The general form of the curves 1–4 suggests that for the lower fields of the order of 5×10^4 v cm^{-1} the graphs tend to a uniform slope. Further, the current depended mainly upon the value of the electric field, irrespective of the gap width, provided that the latter was not so great that the total applied voltage was sufficient to cause a current pulse. For example, the field currents represented by

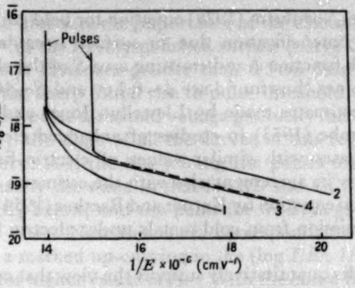


Figure 2. Pre-breakdown currents with the occurrence of pulses

the curves 3 and 4 are practically coincident, and for a given point on curve 4 (the 1·03 cm gap) the voltage is double that for the corresponding point on curve 3. At the larger distances, greater than 1 cm, it was only possible to establish high values of the field E by using correspondingly high values of anode potential V, and such values of V produced successions of pulses. High values of E could, however, be obtained with low V by using much smaller electrode distances. Thus, in order to investigate the field currents over a wider range of values of E without the production of pulses, currents were measured using a gap distance of 0·15 cm; and although values of E up to 80 kv cm^{-1} were used the anode potential was less than 12 kv and no pulses were produced. The full curve involving currents up to 10^{-6} A thus found is given in figure 3. The field E could in this way be increased until the current became large enough to be limited only by the circuit conditions corresponding to the full breakdown of the so-called vacuum. The slope of the linear curve of figure 3 for the 0·15 cm gap, where the values of V extend over a much wider range to much lower values is approximately the same as those in figure 2.

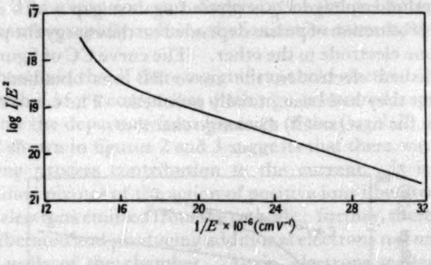


Figure 3. Pre-breakdown currents in the 0·15 cm gap.

These results are consistent with the view that the pre-breakdown currents were due to electron emission from the cold cathode at fields of the order of 10^5 v cm^{-1}, the emission being dependent upon the electric field and not upon the gap voltage. Within the limits of the experiment it is possible to compare the data with the Fowler–Nordheim (1928) equation for field emission, making allowance for local field intensification due to surface irregularities. In this way estimates of the work function ϕ and emitting area S of the electron source can be obtained, and the values thus found are $\phi \sim 0{\cdot}1$ ev and $S \sim 10^{-14}$ cm^2. These are in agreement with estimates made by Llewellyn Jones and Morgan (1953) and Morgan and Harcombe (1953) in studies of enhanced electron emission from cold electrodes in gases with similar values of electric field, of the order of 10^4–10^5 v cm^{-1}, and is in agreement also with the estimates made on the basis of the Fowler–Nordheim equation by Kerner and Raether (1954) from their measurements of electron emission from cold metals under electric fields of the order of 10^4 v cm^{-1}.

The present results quantitatively support the view that cold electron emission (at least, field-dependent emission from electrodes at room temperature) plays a part in initiating breakdown at these low gas pressures of the order of 10^{-5} mm Hg.

§ 5. CURRENT PULSES

The pulse threshold voltage V_{p} was measured for different gap widths and different electrode surface conditions. The pulse threshold voltages were definite and critical. There was no evidence of a change in the relation between the magnitude of the current I and the value of $E\,(=V/d)$ as conditions approached the pulse stage; the occurrence of a pulse was quite sudden. For example, curve 4 (figure 2) is coincident with curve 3 until V reaches about 66 kv when a pulse occurred terminating curve 4. There is no gradual departure from coincidence as the voltage approached this value. The occurrence of a pulse at large values of electrode separation appeared to be mainly determined by the total gap voltage V and the nature of the electrode surface rather than by the value of the electric field. The steady current I was, however, determined by the field V/d irrespective of the separate values of V or d, i.e. these currents were independent of the actual potential of the anode. On the other hand, the pulses seemed to be separate and distinct phenomena from the current I, independent of (or dependent only to a subsidiary extent on) E, but dependent on V. This is illustrated by figure 4 where pulse threshold voltage V_{p} is plotted against gap width d. This at once suggests that the production of pulses depended on the energy the particles acquired in passing from one electrode to the other. The curve CC of figure 4 was obtained with freshly polished electrodes, the curve BB was obtained with electrodes several weeks after they had been initially polished. The curve AA was obtained with electrodes in the first (small) discharge chamber.

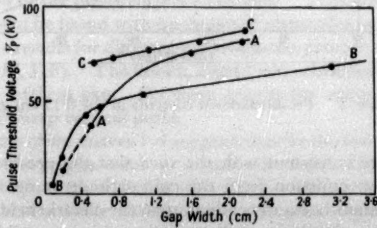


Figure 4. Pulse threshold voltage as a function of electrode separation.

These results also illustrate the importance of the electrode surface conditions in determining V_p: for the freshly polished surface (curve CC), V_p was approximately 25 kv greater at distances greater than 0·5 cm than for the aged surface (curve BB). After the sharp initial rise in pulse threshold voltage shown by the curves AA and BB, the pulse threshold voltage gradually increased with increasing electrode separation; the slope of all the curves in this region is approximately uniform. In the curve AA, the line joining the points for the gaps of 1·6 and 2·16 cm is approximately the same as that for BB and CC, but the point for the 1·03 cm gap is slightly below, and the point for 0·52 cm gap is much below the continuation of this line. Now this curve corresponds to the results shown in figure 3 which have a marked up-curving in the ($\log I/E^2$, $1/E$) graphs of the pre-breakdown current for higher values of E. With the curve BB, on the other hand, when the pre-breakdown currents were smaller and when there appeared to be no significant up-curving in the ($\log I/E^2$, $1/E$) graphs, the (V_p, d) curve continues at the small slope down to much lower values of d ($<0·5$ cm) before V_p becomes markedly reduced. This is also the case for the set of results represented by curve CC and was also noted in many subsequent measurements. A group of results is given in the table, from which it appears that the magnitude of the pre-pulse current exerted some influence in the onset of pulses, provided, however, that the anode potential was adequate.

Gap width d (cm)	0·15	0·52	1·03	1·6	2·15
Pulse threshold V_p (kv)	†	37	66	77·5	83
I (10^{-8} A)	10^2	1	2·5	$2·5.10^{-2}$	$2·5.10^{-3}$
$E(=V/d)$ (kv cm^{-1})	80	71	64	48·5	38·5

† No pulse with $V=12$ kv.

It has been pointed out above that the gas density may be considerable for a short time in the immediate vicinity of the area on the electrodes from which the gas was evolved. Now this evolution of gas at once suggests a reason for the self-extinguishing of the pulse. When the electron density or particle energy at the anode becomes sufficient to cause a momentary liberation of gas, critical ionization of the gas molecules must immediately occur in that region, giving rise to a pulse. The expansion and dispersion of the gas throughout the large volume of the chamber rapidly reduces the ionization current, thus extinguishing the pulse.

It appears, therefore, that if there is significant cold electron emission the pulse threshold, or threshold for a continuous discharge, is lowered, especially at smaller gap widths. Now the departure from linearity of the ($\log I/E^2$, $1/E$) curve at the high values of E shown in figures 2 and 3 suggests that there was some additional current-generating process contribution to the current. It is not difficult to envisage a simplified picture of the action of positive ions liberated at the anode by the high-energy electrons emitted from the cathode; further, there is the possibility of x-rays being liberated and producing additional electrons not only at the cathode but also at the walls of the chamber. Again, electrons scattered at the anode surface can lead to further electron emission from the cathode by causing fluorescence of the glass walls, leading ultimately to cathode photoelectric emission.

It is suggested that the results are consistent with a picture drawn on the following lines. Suppose an initial field emission current I_0 leaves the cathode; let the coefficients of ion- and x-ray photon-generation per incident electron at the anode be A and B respectively, and let f be the absorption coefficient for electrons incident at the anode. Let C be the number of electrons produced per incident ion at the cathode, and D the fraction of x-rays liberated which release an electron from the cathode. Again, suppose that of the $I_0(1-f)$ reflected electrons a fraction E causes fluorescence of the glass walls leading to the liberation of G photoelectrons from the cathode. Then the total gap current may be shown to be given by

$$\begin{aligned} I &= I_0(1+fA)/[1-\{f(AC+BD)+(1-f)EG\}], \\ &= I_0(1+fA)/(1-\xi). \end{aligned}$$

If the denominator approaches zero then the current can increase catastrophically. The condition for instability and rapid increase in current is therefore $\xi \geqslant 1$.

This criterion is fundamentally the same as that given by Trump and Van de Graaff (1947) in their particle exchange theory of vacuum breakdown. If the products AC, BD and EG are sufficiently large such a particle exchange process could lead to the occurrence of pulses and finally to breakdown.

This view is supported by the results obtained with the 0·15 cm gap where the current increased steadily at comparatively low voltage and finally passed into a discharge without an abrupt discontinuity, and where I_0 was appreciable. For gaps greater than 2 cm, pulses occurred without apparently any gradual increase in I_0. However, it must be remembered that the electric field for the larger gaps was much smaller than for the 0·15 cm gap although the total voltage was greater. The present work shows that I_0 decreases exponentially with decreasing field so that I_0 would be small if the voltages were such that the x-rays and photons could be produced with the larger gap widths. However, once the critical voltage for x-ray liberation is reached the products BD and EG might become rapidly appreciable and a pulse occur for small increases of V: there would be no gradual increase of current with V. The fact that the pulse threshold voltage was critical tends further to support this view. Even at smaller distances and reduced voltages the field would be larger and I_0 appreciably bigger, and a more gradual increase in I would take place, since the main contribution would now be due to positive ions released at the anode and secondary electrons released by the ions on impact with the cathode.

The value of the secondary ionization coefficient γ due to H^+ ion impact on Cu has been determined by Hill, Buechner, Clark and Fisk (1939) and for ion energy of 100 kv is about 2–3. Values of the secondary electron coefficient δ are much less than this (Trump and Van de Graaff 1947). The product AB, which may be of the order of magnitude of $\gamma\delta$, is thus small and certainly does not approach sufficiently near unity to cause any marked increase in the total current. It is doubtful whether these two factors A and B can alone lead to breakdown without any contribution due to x-rays and photons unless there is some other process active by which their effect is greatly enhanced. The release of gas from electrodes would provide the medium for such an enhancement by the amplification of the initial current in ionizing collisions with gas molecules. If the gas pressure was maintained, this action would lead to a low pressure gas discharge. It is not unlikely that the continuous discharge noted in § 3 was of this type. Further evidence on the importance of x-rays and photons in the particle exchange process was provided by the observed gradual increase of V_p as the electrodes were separated, as shown in figure 4. At increasing distances the solid angle presented by the cathode diminished and the probability of x-ray interaction with the cathode was correspondingly reduced. To lead to breakdown additional x-rays would then be required as the gap increased. Now the x-ray intensity increases with the square of the voltage, so that raising the voltage would again lead to conditions tending to satisfy the breakdown criterion. This explains qualitatively the gradual slope of the (V_p, d) curve at larger distances and voltages.

§ 6. RESULTS WITH IMPROVED APPARATUS

At the larger electrode distances with the original discharge chamber there was some indication that effects other than those due to electron emission from the electrodes were occurring. Consequently the apparatus was modified, the electrode shape and the screening of leads to the cathode were improved, an enlarged chamber being used to provide greater distances from the electrodes to the glass walls.

With this later improved form of apparatus using freshly polished and cleaned electrodes it was at first found that the pre-pulse currents were too small to be measured accurately, but after some weeks measurable currents were obtained over a much wider range of fields than was previously possible. The points obtained with low fields and those obtained with high fields all fell on the same straight lines which had approximately the same slope as the linear portions of the graphs in figures 2 and 3. At no time did they exhibit marked up-curving portions similar to those graphs. In general the currents were rather smaller than those corresponding to the linear part of the graphs. With gaps of 0·2 cm and using smaller but more stable generators, it was possible to get currents of as much as 10^{-9}A without the occurrence of a pulse.

With the freshly polished electrodes pulses appeared at lower threshold voltages than with the previous apparatus, and a graph of V_p against d was an approximately straight line nearly parallel to the axis, indicating very little dependence on any factor other than the anode potential. This relation held for gap widths between 0·8 cm and 4·5 cm. Below 0·8 cm the threshold voltages were reduced. The experiments showed that the nature of the anode surface had an important effect upon the pulse threshold voltage. The threshold voltage could be reduced by careful polishing and cleaning of the surface, but the pre-pulse field currents from the unaltered cathode remained unchanged.

The same general phenomena of the development of current observed in the preliminary work was again noted; the generation of X-rays and fluorescence of the chamber walls accompanied the occurrence of pulses. Again, as with the original chamber and electrodes, evolution of gas was noted and with a small gap of about 0·15 cm a type of discharge could be obtained in the form of a continuous succession of narrow sparks without the prior occurrence of a pulse. Once initiated, the discharge appeared to be entirely governed by the nature of the external circuit and alteration of the impedance of the latter indicated that the current would increase to high arc values. For the 0·15 cm gap the total anode potential was about 30 kv while the field was over 10^5 v cm^{-1}. The results obtained with the improved apparatus thus confirmed the results obtained with the first chamber.

§ 7. CONCLUSIONS AND SUMMARY

The present work supports previous suggestions that vacuum breakdown is initiated by cold electron emission; however, the order of magnitude of the fields necessary to cause this emission are only about 10^4 to 10^6 v cm^{-1}. Under suitable conditions this initiatory emission can lead to complete breakdown with gas-discharge-like characteristics preceded by current pulses. At larger gap distances and with low electric fields the action of X-rays and photons appear to be necessary, whereas in shorter gaps and higher fields, the relative importance of the action of positive ions is increased, particularly when there is a release of gas. At larger distances breakdown is determined mainly by the total gap voltage, whereas at smaller distances it is determined to some extent by the electric field and magnitude of pre-pulse currents. In both cases the state of the electrode surfaces is important, especially when gas can be evolved under electron bombardment at the anode or ion bombardment in the cathode.

The development of current passing between un-outgassed electrodes with an approximately uniform field at pressures of the order 10^{-5} mm Hg is thus characterized by four stages :

(1) small field-dependent currents $\leqslant 10^{-9}$ A,

(2) current pulses at a critical anode voltage,

(3) larger repetitive pulses at a higher voltage with which the evolution of gas was readily observed,

(4) more complete breakdown with current determined by the circuit constants.

The occurrence of a pulse is a separate process from (1). Over a wide range of gap widths of from 0·5 cm to 4·0 cm the pulse voltage was nearly constant, an eightfold increase in gap width causing a 50% increase in pulse voltage. Hence with a short gap the pre-pulse current can reach a high value before a pulse intervenes. The pulse is quite sudden and no evidence was ever obtained of any build up of current as the conditions for a pulse approached.

The occurrence of pulses involves evolution of gas, primarily due to bombardment of the electrodes. There was, however, no evidence of discharges due to the residual gas. The importance of the walls of the chamber was very evident.

ACKNOWLEDGMENTS

The author wishes to offer his thanks to the Chairman, Sir George Thomson, and members of the Warren Research Committee of the Royal Society for the appointment which enabled him to undertake this work, and to Professor W. John of the Electrical Engineering Department, Queen Mary College, London, for extending the facilities of his laboratories.

REFERENCES

AHEARN, A. J., 1936, *Phys. Rev.*, **50,** 238.
VAN ATTA, L. C., and VAN DE GRAAFF, R. J., 1933, *Phys. Rev.*, **43,** 158.
BENNET, W. H., 1932 a, *Phys. Rev.* **39,** 182 ; 1932 b, *Ibid.*, **40,** 416 ; 1933, *Ibid.*, **44,** 859 ; 1934, *Ibid.*, **45,** 890.
COOLIDGE, W. D., 1913, *Phys. Rev.*, **2,** 409.
CRANBERG, L., 1952, *J. Appl. Phys.*, **23,** 518.
DOLAN, W. W., DYKE, W. P., and TROLAN, J. K., 1953, *Phys. Rev.*, **91,** 1054.
DYKE, W. P., and TROLAN, J. K., 1953, *Phys. Rev.*, **89,** 799.
DYKE, W. P., TROLAN, J. K., DOLAN, W. W., and BARNES, C., 1953 a, *J. Appl. Phys.*, **24,** 270.
DYKE, W. P., TROLAN, J. K., MARTIN, E. E., and BARBOUR, J. P., 1953 b, *Phys. Rev.*, **91,** 1043.
FOWLER, R. H., and NORDHEIM, L., 1928, *Proc. Roy. Soc.* A, **119,** 173.
HILL, A. G., BUECHNER, W. W., CLARK, J. S., and FISK, J. B., 1939, *Phys. Rev.*, **55,** 463.
KERNER, K., and RAETHER, H., 1954, *Z. angew. Phys.*, **6,** 212.
KILPATRICK, W. D., 1953, *U.C.R.L.*, No. 2321 (September).
LANGMUIR, I., 1913, *Phys. Rev.*, **2,** 450.
LLEWELLYN JONES, F., and MORGAN, C. G., 1953, *Proc. Roy. Soc.* A, **218,** 88.
MORGAN, C. G., and HARCOMBE, D., 1953, *Proc. Phys. Soc.* B, **66,** 665.
TRUMP, J. G., and VAN DE GRAAFF, R. J., 1947, *J. Appl. Phys.*, **18,** 327.

Recombination Theory for Indium Antimonide

By P. T. LANDSBERG and T. S. MOSS
Radio Department, Royal Aircraft Establishment, Farnborough, Hants.

MS. received 12*th December* 1955, *and in revised form* 27*th January* 1956

Abstract. The theory of the recombination of holes and electrons via traps is extended to include the effect of degeneracy. Special attention is paid to the example of InSb.

Values for the radiation lifetime for specimens of various carrier concentrations are also calculated.

---

§ 1. Introduction

The object of the present paper is to extend the theory of the recombination of holes and electrons by trapping mechanisms to the case of degenerate semiconductors. This extension is of some intrinsic interest since recombination theory has always been worked out in a semi-classical approximation in which degeneracy effects have been ignored. This applies not only to recombination via traps, but also to radiative recombination between bands (Van Roosbroeck and Shockley 1954) and to recombination by Auger effect (Pincherle 1955). The extended theory is also of practical interest since it is required for comparison with experimental studies on the lifetime of carriers in InSb, though the detailed data available for this comparison are as yet rather inadequate. No attempt is made here to assess which of the various recombination mechanisms dominates in InSb, but merely to discuss quantitatively some of the effects which recombination via traps is likely to have in this material. In germanium recombination via traps can be the dominant mechanism (Burton, Hull, Morin and Severiens 1953).

The potentialities of InSb as a long wavelength photoconductor have already been shown (Moss 1955 a, b), so that a study of the lifetime of injected carriers in this material is of interest. It is also well known that this substance becomes degenerate at comparatively low electron concentrations ($\sim 7\times10^{16}\,\mathrm{cm}^{-3}$). This departure from classical statistics can in fact be seen very clearly by plotting the ratio np/n_i^2 as a function of the position of the Fermi level (figure 1), where n, p are respectively the densities of electrons in the conduction band and holes in the full band, and n_i^2 is the value which the product np would have if the sample were truly non-degenerate, so that the departure of np/n_i^2 from unity is a measure of the 'degeneracy' of the sample. In figure 1, and throughout the rest of this paper, it is assumed that the temperature of the material is 290°K, that the main energy gap in InSb is 0·18 ev at this temperature, and that the effective mass ratios m_e/m, m_h/m are respectively 0·03 and 0·23.† The position of the Fermi level E_F is measured by the energy $E_F - E_V$ which separates it from the top of the valence band E_V, expressed in units of the main energy gap $E_g(=0{\cdot}18\,\text{ev})$. With the above values n_i is found to be $1{\cdot}56 \times 10^{16}\,\text{cm}^{-3}$.

---

† Departures from these values given by low temperature cyclotron resonance experiments are not expected to affect the order of magnitude of the results obtained in this paper.

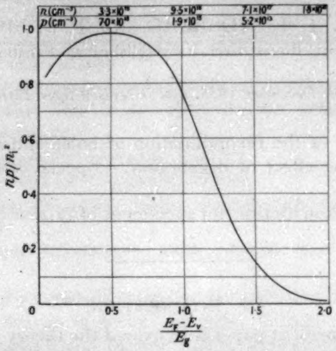


Figure 1. The departure of np/n_i^2 from unity as a measure of 'degeneracy' in InSb. The position of the Fermi level is plotted horizontally and depends on the composition of the sample.

§ 2. Recombination Theory

In this section the theory will be worked out in general terms, application to InSb being postponed to § 4. Three groups of quantum states will be considered. Quantities pertaining to the conduction states in the almost empty band will be given suffixes e, quantities pertaining to the conduction states in the almost full band will be given suffixes h, and quantities which refer to localized levels are given suffixes j $(=1, 2, \ldots)$, where different values of j distinguish states belonging to different energies. The number of states of type j per unit volume is G_j. The number of conduction states per unit volume in an energy range dE, on the other hand, is denoted by $N(E)\,dE$. Let A_{ej}, A_{hj} be the probabilities per unit time that an electron or hole respectively makes a transition from a conduction state to a localized state, and A_{je}, A_{jh} be the probabilities per unit time for the converse processes. The net rate of capture of conduction band electrons, which lie in the energy range dE_e, by traps of type j is then

$$dU_{ej} = N_e G_j [A_{ej} P_e (1 - P_j) - A_{je} P_j (1 - P_e)]\, dE_e, \qquad \ldots\ldots(1)$$

where P_e, P_h denote occupation probabilities of a conduction state by electrons and holes respectively, and are given by Fermi functions. P_j is the occupation probability of a localized level and is of the form

$$P_j = \frac{1}{1 + \beta_j \exp(\eta_j - F_j)}, \qquad \ldots\ldots(2)$$

where η_j represents the energy of the localized level, i.e. $E_j/\boldsymbol{k}T$, F_j is the quasi Fermi level of the localized state, also in units of $\boldsymbol{k}T$, and β_j depends on the nature of the trapping atom (Landsberg 1952). This may be $\frac{1}{2}$, 2 or a more complicated function, which may be temperature dependent. The need to introduce the

β's explicitly may be removed by the convention, which will here be made, that 'effective' energies are used for localized levels; these have the form

$$E_j + \boldsymbol{k}T \log \beta_j. \qquad \ldots\ldots(3)$$

Hence, making use of the fact that dU_{ej} vanishes at equilibrium to eliminate A_{je}/A_{ej},

$$dU_{ej} = P_e N_e A_{ej} G_j (1-P_j)[1-\exp(F_j - F_e)]\, dE_e. \qquad \ldots\ldots(4)$$

In equilibrium the quasi Fermi levels (divided by $\boldsymbol{k}T$) for the localized levels F_j and the conduction band F_e are, of course, equal. Defining $\overline{A_{ej}}$ as

$$(1/n)\int P_e N_e A_{ej}\, d\eta_e,$$

and similarly for A_{je}, one finds after integrating over the conduction band,

$$U_{ej} = n[1-\exp(F_j - F_e)](1-P_j)G_j\overline{A}_{ej}. \qquad \ldots\ldots(5)$$

A similar argument for holes leads to

$$U_{hj} = p[1-\exp(F_h - F_j)]P_j G_j \overline{A}_{hj} \qquad \ldots\ldots(6)$$

for the net rate of capture of holes by traps of type j.

The state of the system which is of interest is one in which the transition rates (5), (6) are non-zero and constant in time, and in which the electron–hole recombination rate, which they represent, is balanced exactly by an electron–hole generation rate which is induced by external sources (e.g. incident light). In such a steady non-equilibrium state the P_j's must also be constant in time, i.e. $U_{ej} = U_{hj}$. Electron transfer between traps of different types and direct transitions between conduction and valence bands are neglected. By equating (5) and (6), then, one obtains an equation for P_j, which can therefore be eliminated from these equations. Hence, the steady state recombination rate via traps of type j, which will be denoted by U_j, is found to be

$$U_j = \frac{\overline{A_{ej}}\,\overline{A_{hj}}\,G_j np[1-\exp(F_h - F_e)]}{\overline{A_{ej}}n[1+\exp(\eta_j - F_e)] + \overline{A_{hj}}p[1+\exp(F_h - \eta_j)]}. \qquad \ldots\ldots(7)$$

The total recombination rate U due to trapping is obtained by summing (7) over all levels j.

It will be supposed that the steady state conditions envisaged represent small departures from equilibrium. Giving equilibrium quantities an additional suffix zero, we have

$$n_0 = N_0 G_{1/2}(F_0 - \eta_c),\ p_0 = P_0 G_{1/2}(\eta_v - F_0),$$

where $N_0 \equiv 4\pi \boldsymbol{h}^{-3}(2m_e\boldsymbol{k}T)^{3/2}$, and P_0 is the corresponding expression for holes; n_c, η_v are the energies (divided by $\boldsymbol{k}T$) at the bottom of the conduction band and the top of the valence band respectively. Also

$$G_c(y) \equiv \int_0^\infty \frac{x^c\,dx}{1+\exp(x-y)}.$$

Hence

$$\delta n = n_0 \alpha_e \delta F_e, \quad \delta p = p_0 \alpha_h \delta F_h \qquad \ldots\ldots(8)$$

where

$$2\alpha_c \equiv \frac{G_{-1/2}(F_0 - n_c)}{G_{1/2}(F_0 - \eta_c)}, \quad 2\alpha_h \equiv \frac{G_{-1/2}(\eta_v - F_0)}{G_{1/2}(\eta_v - F_0)}. \qquad \ldots\ldots(9)$$

n in (7) is now interpreted as $n_0 + \delta n$ and similarly with p, F_e, F_h, subject to the assumption $\delta n = \delta p$.

Defining the lifetime τ of the additional carriers as $\delta n/U$, τ can be obtained by reciprocal addition of all the lifetimes $\tau_j \equiv \delta n/U_j$, associated with the individual trapping states. Using (7) and (8), it is found that

$$\tau_j = \frac{1}{G_j \overline{A}_{hj}} \frac{n_0 \alpha_e \alpha_h}{n_0 \alpha_e + p_0 \alpha_h} [1 + \exp(\eta_j - F_0)] \left[1 + \frac{\overline{A}_{hj} p_0}{\overline{A}_{ej} n_0} \exp(F_0 - \eta_j) \right] \quad \ldots\ldots(10)$$

which is the desired lifetime. It can readily be shown to go over into equation (5.5) of Shockley and Read (1952) if degeneracy is neglected.

It is convenient to replace the second factor in (10) by $n_0/(n_0+p_0)$. This increases (10) by a factor which always lies between 1 and 1·05 in the case of InSb. It is also a very good approximation in most other semiconductors. This is due to the fact that if the Fermi level lies low in the forbidden band, and $\alpha_h < 1$, $\alpha_e \sim 1$, the electron population is negligible compared with the hole population, and α_h cancels. On the other hand, if the Fermi level lies high in the forbidden band, and $\alpha_h \sim 1$, $\alpha_e < 1$, the hole population is negligible compared with the electron population, and α_e cancels. In either case the α which deviates from unity tends to cancel.

§ 3. DEPENDENCE OF LIFETIME ON COMPOSITION (FERMI LEVEL)

It will be supposed in this section that lifetimes have been measured at constant temperature for a range of impurity concentrations on a material whose effective masses, energy gap and impurity levels are reasonably well known, so that the position of the Fermi level can be estimated. It will also be supposed that the dominant recombination mechanism for at least part of the curve of lifetime against Fermi level is the trapping by a single type of centre. The first point to be observed in this case is that the position of the maximum of this curve can sometimes be used to suggest which of the known trapping levels may be responsible. This will now be demonstrated.

The following simplified form of (10) will be used:

$$G_j \overline{A}_{hj} \overline{A}_{ej} \tau_j = (n_0+p_0)^{-1} [\overline{A}_{ej} n_0 + \overline{A}_{hj} p_0 \exp(F_0 - \eta_j)][1 + \exp(\eta_j - F_0)]. \quad \ldots(11)$$

Differentiating with respect to F_0, and equating to zero, we find, after some computation, the following condition for a maximum:

$$\frac{\mu\alpha_e + \lambda_j \exp(F_0 - \eta_j)}{\mu + \lambda_j \exp(F_0 - \eta_j)} - \frac{\mu\alpha_{e0} + 1}{\mu + 1} + \frac{1 + \alpha_h}{\mu + 1} = \frac{1}{1 + \exp(F_0 - \eta_j)} + \frac{\alpha_h}{1 + \mu\lambda_j^{-1} \exp(\eta_j - F_0)}, \quad \ldots\ldots(12)$$

where $\lambda_j = \overline{A}_{hj}/\overline{A}_{ej}$ and $\mu = n_0/p_0$. ......(13)

If the maximum occurs when the Fermi level is not too close to either edge of the forbidden gap, it will be a very good approximation to simplify (12) by putting

$$\alpha_e = \alpha_h = 1, \quad \mu = (m_e/m_h)^{3/2} \exp(2F_0 - \eta_c - \eta_v). \quad \ldots\ldots(14)$$

We shall write F_{jm} for that equilibrium value of the Fermi level for which a maximum occurs if trapping is by the jth type of centre. The resulting form of equation (12) is quadratic in $\exp(F_{jm})$, and its solution can be written in the form

$$\exp(F_{jm}) = -w_j + (v^{-1} + w_j^2)^{1/2}, \quad \ldots\ldots(15)$$

where

$$v \equiv (m_e/m_h)^{3/2} \exp(-\eta_c - \eta_v), \quad \ldots\ldots(16a)$$

$$w_j \equiv \lambda_j^{-1} [v \exp \eta_j + \lambda_j \exp(-\eta_j)]^{-1} \quad \ldots\ldots(16b)$$

To show that $\exp F_{jm}$, as given by (15) and (16), satisfies two inequalities, note first that, by (16*b*) and since λ_j is positive,

$$-v^{-1}\exp(-\eta_j)<w_j<\exp\eta_j. \qquad \ldots\ldots(17)$$

Note next that

$$\frac{d(\exp F_{jm})}{dw_j}=\frac{w_j}{(v^{-1}+w_j^{2})^{1/2}}-1<0. \qquad \ldots\ldots(18)$$

By (17) and (18) the least value of w_j corresponds to the largest value of F_{jm}, and conversely, so that

$$-\exp\eta_j+(v^{-1}+\exp 2\eta_j)^{1/2}<\exp F_{jm}<v^{-1}\exp(-\eta_j)+[v^{-1}+v^{-2}\exp(-2\eta_j)]^{1/2}. \qquad \ldots\ldots(19)$$

Thus with each type of trapping centre there is associated a range of positions of the Fermi level such that the corresponding lifetime can have a maximum only if the Fermi level lies within that range. It is therefore possible to use the experimental lifetime to determine which of the possible trapping levels may be responsible.

Once a likely dominant trapping level has been determined, the validity of the present theory may be tested by using the experimentally determined relationship between lifetime and Fermi level to plot

$$y\equiv\frac{\tau_j}{1+\exp(\eta_j+F_0)}\frac{n_0+p_0}{n_0}\quad\text{against}\quad x\equiv\frac{p_0}{n_0}\exp(F_0-\eta_j). \qquad \ldots\ldots(20)$$

Equations (10) or (11) lead one to expect the straight line

$$y=(G_j\overline{A_{hj}})^{-1}(1+\lambda_j x). \qquad \ldots\ldots(21)$$

Confirmation of (21) would lend support to the theory, and would lead to an empirical estimate of the ratios λ_j of the transition probabilities. If a straight line is not obtained, other traps or other recombination mechanisms must play an important part.

The following approximations are also useful:

$$G_j\overline{A_{hj}}\tau_j\simeq\begin{cases}\lambda_j+(N_0/p_0)\exp(\eta_j-\eta_c) & (F_0<\eta_j,\ n_0\ll p_0), \qquad \ldots\ldots(22)\\ 1+\lambda_j(P_0/p_0)\exp(\eta_V-\eta_j) & (F_0>\eta_j,\ n_0\gg p_0). \qquad \ldots\ldots(23)\end{cases}$$

§ 4. APPLICATION TO InSb

Using the constants given in § 1 we can now calculate the lifetime τ_1 when there is a trapping level whose effective energy is at the bottom of the conduction band, and similarly a lifetime τ_2 for a level at the top of the valence band, i.e. $\eta_1=\eta_c$, $\eta_2=\eta_V$. Typical computed curves for τ_1 are shown in figure 2 and for τ_2 in figure 3. τ_{01} is $(G_1\overline{A_{h1}})^{-1}$ and τ_{02} is $(G_2\overline{A_{h2}})^{-1}$.

Suppose now that in plotting the curve given by (11) one were to compute the carrier densities not by using the Fermi integrals, but by using the classical approximation $G_{1/2}(y)\sim(\pi^{1/2}/2)\exp y$ to these integrals. This is equivalent to assuming the semiconductor to be non-degenerate. This approximation must in the present case be expected to affect the right-hand sides of the curves of figures 2 and 3. In the case of τ_1 the error produced lies within 1% for $\lambda_1\sim 1$. On the other hand, τ_2 would be markedly depressed below the values shown in figure 3 for values of $(E_F-E_V)/E_g$ lying between 1 and 3. For $\lambda_2=1$ the error rises to 64% at $(E_F-E_V)/E_g=1{\cdot}5$. This difference may be understood by reference to equation (23). The departure from the limiting value $(G_j\overline{A_{hj}})^{-1}$ of the lifetime is determined by the factor $(n_0 \exp \eta_j)^{-1}$, and is therefore largest for the lower lying levels. Also, an approximation which overestimates n_0 underestimates this departure from the limiting value. This is what happens if the classical approximation is used for these high lying Fermi levels.

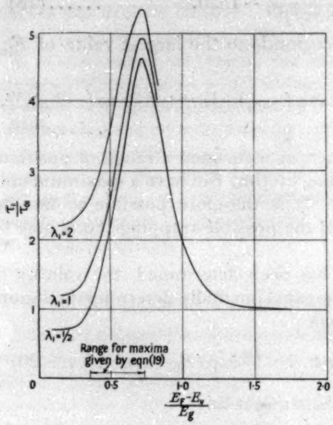


Figure 2. The lifetime of holes and electrons in InSb if they recombine only via traps whose effective energy level lies at the bottom of the conduction band. Other assumptions as discussed in the text.

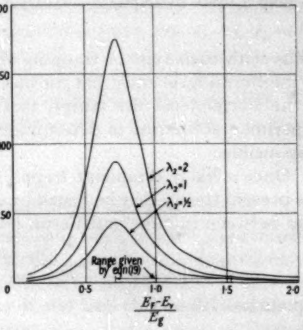


Figure 3. The lifetime of holes and electrons in InSb if they recombine only via traps whose effective energy level lies at the top of the valence band. Other assumptions as discussed in the text.

The ranges available for the Fermi level at which the lifetime is a maximum must now be estimated. We have

$$v^{-1} = (m_h/m_e)^{3/2} \exp \eta_c = 21{\cdot}2 \exp 7{\cdot}2 \sim 28\,200,$$

where the top of the valence band has been used as the energy zero. Hence (19) yields $2{\cdot}56 < F_{1m} < 5{\cdot}25$, $5{\cdot}12 < F_{2m} < 10{\cdot}94$. On the horizontal axis of figures 2 and 3 these ranges correspond to the intervals (0·36, 0·73) and (0·71, 1·52). Preliminary experimental results suggest a maximum in the lifetime somewhere in the range (0·6, 0·8), so that it is not yet possible to point to either trapping level as probably carrying the main recombination traffic.

It is easy to show from (16*b*) that

$$\frac{dw_j}{d\lambda_j} = \frac{v \exp \eta_j + \exp(-\eta_j)}{[\lambda_j \exp(-\eta_j) + v \exp \eta_j]^2} > 0. \qquad \ldots\ldots(24)$$

From (18) and (24)

$$\frac{d(\exp F_{jm})}{d\lambda_j} < 0. \qquad \ldots\ldots(25)$$

In the case of the curves of figures 2 and 3 the Fermi level does in fact decrease as λ is increased, but the change is very slight, as is clear from the table. All values given in the table correspond to values of $(E_F - E_V)/E_g$ which lie near 0·71 or 0·72. It is easy to verify, however, that the maxima of τ_1 and τ_2 separate if λ_1 is increased above 2 and λ_2 is decreased below $\frac{1}{2}$.

Calculated Values of $\exp F_{jm}$ for maximum Lifetimes

	$\lambda=\frac{1}{2}$	$\lambda=1$	$\lambda=2$
1	178	168	159
2	169	168	167·5

§ 5. ESTIMATION OF THE ORDER OF MAGNITUDE OF $\lambda = \overline{A_{hj}}/\overline{A_{ej}}$

Consider a specimen with a trapping level at the bottom of the conduction band. For intrinsic InSb where the Fermi level lies approximately $2\boldsymbol{k}T$ below the conduction band the occupancy of this trap will be $P_j = 0{\cdot}12$. The frequency at which electrons go into and out of these traps will be high, and as a consequence the rates in and out will be much greater than the net rate U_{ej} so that from equation (1) $A_{ej}P_e(1-P_j) \simeq A_{je}P_j(1-P_e)$.

From Mott and Gurney (1948) the average time in the trap is

$$1/\tau_T \sim \nu \exp(-W/\boldsymbol{k}T)$$

where W is the trap depth and $\nu\,(\sim 10^{12}\,\mathrm{c/s})$ is the frequency of the lattice vibrations. We have said $W \simeq 0$ so $\tau_T \sim 10^{-12}$ sec. Now the average number of electrons in the traps P_jG_j will be given by the product of τ_T and the rate of entering the traps $A_{ej}G_j(1-P_j)\int N_eP_e\,dE_e$. Hence

$$A_{ej} \sim \frac{P_j}{1-P_j}(\tau_T n_0)^{-1} \sim 10^{-5} \text{ as } n_0 = 1{\cdot}6\times10^{16}\,\mathrm{cm}^{-3}$$

for intrinsic InSb.

A_{hj} may be estimated from equation (11) using measured values of τ_j. We shall use the value $\tau_j = 4\times10^{-7}$ sec (Moss 1955a), and assume that in this particular sample the main recombination was via traps at the bottom of the conduction band. If any other recombination mechanism were operating in this sample then the lifetime due to trapping would of course be larger, and consequently the values of A_{hj} deduced below would be even smaller.

From (11),

$$G_j\tau_j = [1+\exp(\eta_j - F_0)][\overline{A_{hj}}^{-1} + \overline{A_{ej}}^{-1}\mu^{-1}\exp(F_0-\eta_j)](1+\mu)^{-1}.$$

The value of G_j is given by $1-P_j \sim n_0$ or $G_j \sim n_0$. For all except very p-type samples the $\overline{A_{ej}}$ term may be neglected, and the value of $\overline{A_{hj}}$ is found to decrease progressively from about 10^{-8} at $p_0 = 10^{17}$ to about 10^{-14} at $n_0 = 7\times10^{17}$. It is thus clear that for a wide range of Fermi levels $\lambda \ll 1$.

Hence if the only trapping level present is near the bottom of the conduction band the Fermi level for maximum τ_j will lie near the maximum of the permitted range shown in figure 2, i.e. near 0·7 of the way up the forbidden zone, or approximately $2\boldsymbol{k}T$ below the conduction band, which is the condition for intrinsic material. Conversely, for a trapping level at the top of the full band it would be expected that $\lambda \gg 1$, giving the Fermi level for maximum τ_j near the minimum of the permitted range in figure 3, again corresponding to near intrinsic material.

It is thus tentatively concluded that for low energy traps, whether donors or acceptors, maximum τ_j will occur for approximately intrinsic material, and only if the traps are deeper will the position of maximum τ_j serve to identify the position of the trapping level.

§ 6. RADIATIVE RECOMBINATION

In an uncooled photoconductor, room temperature radiation within the material produces photoelectrons and photo-holes, and by the principle of detailed balancing there must be a complementary process whereby electrons and holes recombine with the generation of radiation (Van Roosbroeck and Shockley 1954). Calculation of this rate of photo-excitation thus gives the radiative recombination rate.

The photo-excitation rate is $R=\int c' q_\lambda K_\lambda \, d\lambda$ where q_λ is the photodensity in the material, K_λ the absorption constant, and c' the radiation velocity in the material. As the significant part of the integral is situated near the long wavelength absorption edge where the material is non-dispersive, c' is constant. We have

$$R=4\left(\frac{c}{c'}\right)^2 \int K_\lambda Q_\lambda \, d\lambda$$

where Q_λ is the tabulated Planck function for radiation *in vacuo*,

$$Q_\lambda = 2\pi c \lambda^{-4} (\exp \boldsymbol{h}\nu/\boldsymbol{k}T - 1)^{-1}.$$

For a specimen with carrier concentrations n, p, and equilibrium values n_0, p_0 the recombination rate

$$R_c = R(np/n_0 p_0), \qquad \ldots\ldots (26)$$

where R is the equilibrium rate (Van Roosbroeck and Shockley 1954). If n and p are small increases, i.e. $n=n_0+\delta n$ and $p=p_0+\delta p$, then

$$R_c = \frac{p\delta n + n\delta p}{n_0 p_0} R = \delta n \frac{p+n}{p_0 n_0} R \text{ and } \tau_R = \frac{\delta n}{\delta R_c} = \frac{n_0 p_0}{n_0 + p_0} R^{-1}. \qquad \ldots\ldots (27)$$

Values of R have been computed for specimens of various carrier concentrations making allowance for the shift of the absorption edge with concentration. The data given by Avery, Goodwin, Lawson and Moss (1954) were used for the complete absorption spectrum for impure InSb, with interpolation between these data and transmission measurements by various workers in the region of the absorption edge for the other specimens. In all cases the extrapolated free carrier absorption was subtracted from the measured data. For the n-type samples this varies as λ^2 (Avery *et al.* 1954).

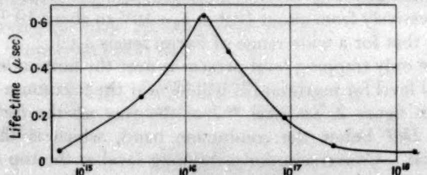


Figure 4. Calculated lifetime for InSb for radiative recombination only.

The corresponding values of τ_R are plotted against electron concentration in figure 4. Results published by Mackintosh and Allen (1955) since this paper was written have shown that the effects of degeneracy on lifetimes is negligible. The lifetimes calculated by these workers agree with those of figure 4.

It will be seen that the radiation lifetime has a maximum value $0{\cdot}6\,\mu$sec for an intrinsic sample. There is uncertainty in this value due to the difficulties in interpolating the absorption data. Errors of as much as 2 : 1 either way are conceivable. Nervetheless the fact that the calculated maximum lifetime is little greater than some measured values of τ (Moss 1955a, b, Avery and Jenkins 1955) is a strong indication that radiative recombination is of importance in InSb.

ACKNOWLEDGMENTS

This work was done while one of us (P. T. L.) was a vacation consultant at the Royal Aircraft Establishment. Acknowledgment is due to the Chief Scientist, Ministry of Supply, for permission to publish the paper.

REFERENCES

AVERY, D. G., and JENKINS, D. P., 1955, *J. Electronics*, **1,** 145.
AVERY, D. G., GOODWIN, D. W., LAWSON, W. D., and MOSS, T. S., 1954, *Proc. Phys. Soc.* B, **67,** 761.
BURTON, J. A., HULL, G. W., MORIN, F. J., and SEVERIENS, J. C., 1953, *J. Phys. Chem.*, **57,** 853.
LANDSBERG, P. T., 1952, *Proc. Phys. Soc.* A, **65,** 604.
MACKINTOSH, I. M., and ALLEN, J. W., 1955, *Proc. Phys. Soc.* B, **68,** 985.
MOSS, T. S., 1955 a, *Proceedings of Atlantic City Conference on Photoconductivity*, 1954 (New York; John Wiley); 1955 b, *J. Electronics*, **1,** 126.
MOTT, N. F., and GURNEY, R. W., 1948, *Electronic Processes in Ionic Crystals* (Oxford : University Press).
PINCHERLE, L., 1955, *Proc. Phys. Soc.* B, **68,** 319.
SHOCKLEY, W., and READ, W. T., 1952, *Phys. Rev.*, **87,** 835.
VAN ROOSBROECK, W., and SHOCKLEY, W., 1954, *Phys. Rev.*, **94,** 1558.

Optical and Electrical Properties of some Oxygen Activated Alkali Halides

By J. EWLES and D. S. BARMBY†

Department of Physics, University of Leeds

MS. received 9th August 1955 *and in amended form* 29th *September* 1955

Abstract. Phosphors showing well-resolved bands in their emission spectra have been made by thermal treatment of alkali halides in the presence of oxygen. The emission spectra are strikingly similar with quite different halides all showing a wave number difference of about 1000 cm^{-1} between the peaks of the main components of the band system. The chlorides and bromides show abnormal thermoelectric effects. It is suggested that both the similarity in the emission spectra and the abnormal thermoelectric effects can be accounted for by the adsorption of O_2^+ on active sites.

§ 1. Introduction

It was observed by one of us some time ago that many white solids became markedly luminescent under ultra-violet light after flaming with a small gas jet. Further investigations showed that in the case of alkali halides the luminescence was of two kinds, one, a broad unresolved band mainly in the blue and violet and the other a band progression of narrow components mainly in the red, yellow and green. The broad band is obtained by rapid heating and cooling by any method, *in vacuo* as well as in air, and is absent with slow heating and cooling. Similar effects are produced by mechanical strain and by the introduction of divalent positive ions. A report of investigations on these effects is given in another paper where it is shown that they are associated with positive ion vacancies. The system of narrow bands which appear on flaming appear also if the molten halide is dropped into liquid air or liquid oxygen but not with liquid nitrogen. They also appear if the halide is heated with any oxygen containing salt. This paper reports investigations of some luminescent and electrical properties of these phosphors and attributes the effects to adsorbed oxygen.

§ 2. Preparation of the Phosphors

2.1. *Materials*

For the alkali halides both Analar and 'luminescent pure' (i.e. showing no luminescence on slow cooling) materials were used. The luminescent spectra produced by dropping into liquid oxygen were the same in both cases. The oxygen containing salts could be any such as peroxide, hydroxide, nitrate, chlorate, etc., and were added up to about one per cent by weight. Neither the nature of the added salt nor its concentration within this limit affected the results.

---

† Now with Wool Industries Research Association, Torridon, Leeds.

2.2. *Activation*

(*a*) *Flaming*. A layer of halide 2–3 mm thick was placed on a platinum tray and a small gas jet played on it till the solid melted.

(*b*) *By liquid air or oxygen*. The halide was melted in a platinum crucible and poured while molten into liquid oxygen or liquid air.

(*c*) *By heating with oxygen containing salts*. The halide was ground into intimate mixture with the oxygen containing salt and then the mixture heated to melting in a platinum crucible and cooled slowly.

§ 3. LUMINESCENT SPECTRA

Emission spectra excited by ultra-violet light from a high pressure mercury vapour lamp were obtained at room temperature, 77°K and 20°K by arrangements previously described (Ewles and Curry 1950). Some of these spectra are shown in the Plate, and the wave numbers of the peaks of the components of the bands in the table. Excitation spectra were obtained by the previously described method of photoelectric fluorescent photometry (Ewles 1949) and are shown in figures 1 and 2.

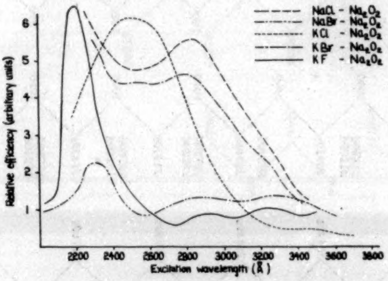


Figure 1. Excitation spectra, various halides.

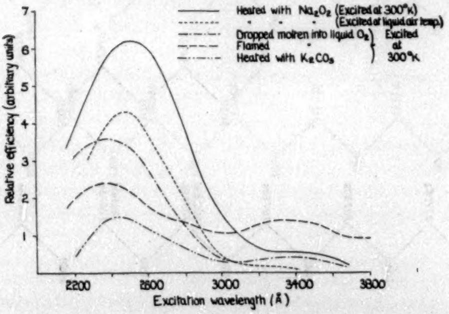
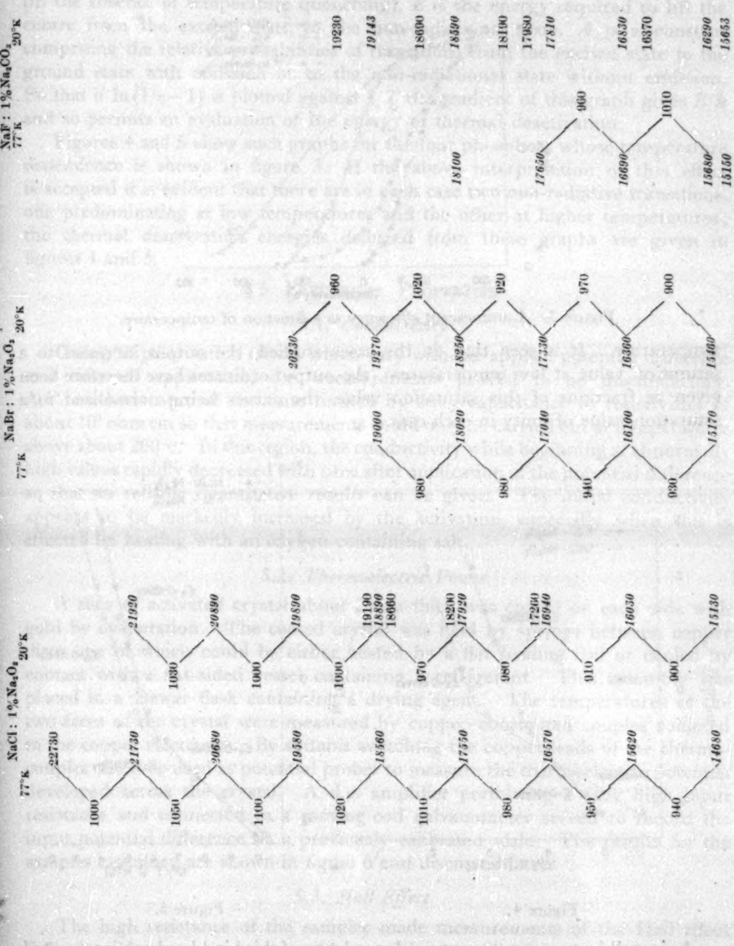


Figure 2. Excitation spectra of KCl activated in various ways.

Wave Numbers of Peaks of Luminescent Emission Band Components

values in cm $^{-1}$ ($\pm$ 30 cm $^{-1}$); d = diffuse, s = sharp; strong components in italics

KCl : 1% Na_2O_2		KCl pure flamed		KCl pure, dropped molten into liquid oxygen				KBr : 1% Na_2O_2				KF : 1% K_2CO_3	
77°K		77°K		77°K		20°K		77°K		20°K		77°K	
	22820s									22610s			
1030											940		
						22340d							
	21790s				*21870s*	*21970s*				*21670s*			
1040				1030		21890d 21200d	1060				1060		
	20750s				*20840s*	*20910s*				*20610s*			
1040				1030			1050				980		
						20200d							
	19710s		*19960*		*19810s*	*19860s*			*19470s*	*19630s*		*19700*	
1030		1010		1050		19720d 19220d	1010	1000			970		1050
	19230d *18680s*		*18950*		*18760s*	*18850s*			*18470s*	*18660s*		*18650*	
1010		970		1000		18760d 18220d	1000	1020			1030		1020
	18160d *17670s*		*17980*		*17760s*	*17850s*			*17450s*	*17630s*		*17630*	
1020		1010		1020		17720d 17240d	1010	970			940		990
	16650s		*16970*		*16740s*	*16840s*			*16480s*	*16690s*		*16640*	
960		970		980		16740d 16260d	970	990			910		930
	16210d *15690s*		*16000*		*15760s*	*15870s*			*15490s*	*15780s*		*15710*	
1040		1000		970		15770d 15230d	1040	930			880		970
	15210d *14650s*		*15000*		*14790s*	*14830s*			*14560s*	*14900*		*14740*	


NaCl : 1% Na_2O_2
77°K
20°K
22730
1000
21730
1050
20680
1100
19580
1020
18560
1010
17550
980
16570
950
15620
940
14680
21920
1030
20890
1000
19890
1000
19190
18890
18660
970
18200
17920
980
17260
16940
910
16030
900
15130
NaBr : 1% Na_2O_2
77°K
20°K
19000
980
18020
980
17040
940
16100
930
15170
20230
960
19270
1020
18250
920
17330
970
16360
900
15460
NaF : 1% Na_2CO_3
77°K
20°K
18100
17650
960
16690
1010
15680
15150
10230
19143
18600
18390
18190
17980
17810
16830
16370
16290
15653

§ 4. Dependence of Luminescent Efficiency on Temperature

The efficiency was compared with that of a target material of constant quantum efficiency (aesculin solution). The luminescent output relative to the excitation energy was measured photoelectrically over a range of temperatures from 183°c to 300°c using the arrangements previously described (Ewles and Lee 1953). In figure 3 is shown the way in which the relative output varies with temperature. It is seen that, in the cases studied, the output increases to a saturation value at low temperatures; the output ordinates have therefore been given as fractions of this saturation value, the curves being normalized to a saturation value of unity in each case.

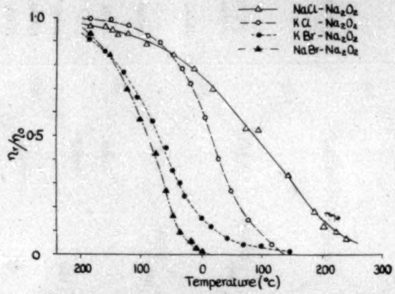


Figure 3. Luminescent efficiency as a function of temperature.

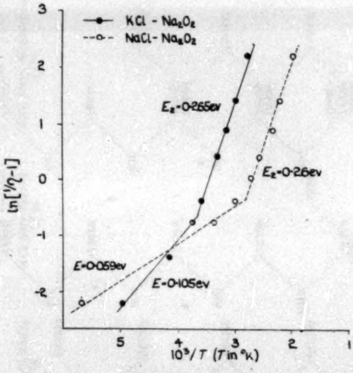


Figure 4.

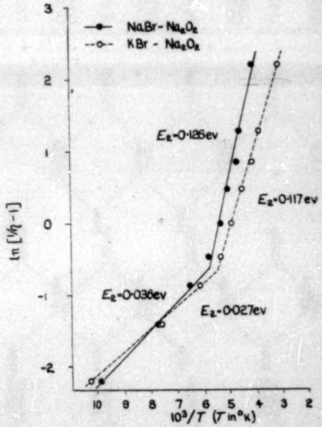


Figure 5.

As is well known, a temperature dependence of this kind has been interpreted by supposing the existence of a non-radiational state from which transitions may occur to the unexcited state without emission of radiation and to which the excited centre may be lifted by thermal energy (von Hippel 1936, Mott 1938). From this idea it is easy to show that

$$\eta_T/\eta_0=\eta=1/[1+A\exp(-E/\boldsymbol{k}T)]$$

where η_T is the efficiency at temperature T, η_0 is the efficiency at saturation (in the absence of temperature quenching), E is the energy required to lift the centre from the excited state to the non-radiational state, A is a constant comprising the relative probabilities of transitions from the excited state to the ground state with emission or to the non-radiational state without emission. So that if $\ln(1/\eta-1)$ is plotted against $1/T$ the gradient of this graph gives $E/\boldsymbol{k}$ and so permits an evaluation of the energy of thermal deactivation.

Figures 4 and 5 show such graphs for the four phosphors whose temperature dependence is shown in figure 3. If the above interpretation of this effect is accepted it is evident that there are in each case two non-radiative transitions, one predominating at low temperatures and the other at higher temperatures; the thermal deactivation energies deduced from these graphs are given in figures 4 and 5.

§ 5. ELECTRICAL PROPERTIES

5.1. *Conductivity*

Owing to polarization effects which vary with the applied potential difference and time of application, d.c. measurements proved to be unsatisfactory. A.C. methods were at the time limited by stray capacitance to resistivities of about 10^7 ohm cm so that measurements could only be carried out at temperatures above about 200°c. In this region, the conductivity while beginning at abnormally high values rapidly decreased with time after application of the potential difference so that no reliable quantitative results can be given. The initial conductivity appears to be markedly increased by the activation, especially when this is effected by heating with an oxygen-containing salt.

5.2. *Thermoelectric Power*

A slice of activated crystal about 2 mm thick was coated on each side with gold by evaporation. The coated crystal was held by springs between copper discs one of which could be either heated by a flat heating coil or cooled by contact with a flat-sided vessel containing a refrigerant. This assembly was placed in a Dewar flask containing a drying agent. The temperatures of the two faces of the crystal were measured by copper–constantan couples soldered to the copper electrodes. By suitable switching the copper leads of the thermocouples could be used as potential probes to measure the thermoelectric potential developed across the crystal. A d.c. amplifier permitting a very high input resistance and connected to a moving coil galvanometer served to record the input potential difference on a previously calibrated scale. The results for the samples examined are shown in figure 6 and discussed later.

5.3. *Hall Effect*

The high resistance of the samples made measurements of the Hall effect difficult. However, it was possible to make a qualitative observation on an activated crystal of KCl. With a field of about 12000 oersteds and a d.c. potential difference of 100 volts a definite Hall effect of negative sign was detected with a very sensitive moving coil galvanometer. Spurious thermoelectric or contact effects were eliminated by reversing the field.

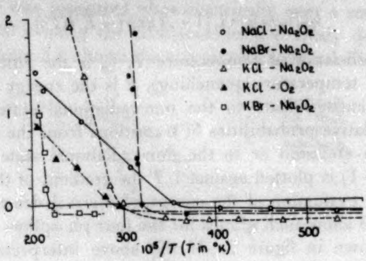


Figure 6.

§ 6. Discussion

The striking similarity of the emission spectra of the phosphors produced with quite different alkali halides dropped when molten into liquid oxygen, heated with an oxysalt or flamed with a gas jet containing oxygen, indicates a common luminescent centre. Honrath (1937) obtained similar phosphors by heating alkali halides with salts containing oxygen or by heating alkali chlorates or bromates or by heating alkali halide crystals in a high pressure of oxygen. He attributed the emission to complexes of the type ClO_4. The nearly equal wave number differences (of about 1000 cm^{-1}) which appear between the components in the emission spectra of halides differing both in anion and cation make it difficult to accept Honrath's suggestion, for it is inconceivable that quite different complexes such as ClO_4 and BrO_4 in quite different crystal fields should have the same vibrational frequency. It seems certain that the presence of oxygen and a somewhat violent thermal treatment in its presence is necessary. It is therefore suggested that the centres are in fact oxygen itself and that the character of the spectra suggest molecular oxygen. The method of preparation indicates too that the formation of fresh crystal surfaces is necessary, suggesting that the oxygen is adsorbed on active sites on these surfaces.† The only known spectrum of molecular oxygen that seems to resemble in position and vibrational frequency the bands reported here is that of O_2^+, 1st negative, in which the transitions are $^4\Sigma_g^- \rightarrow {}^4\Pi_u$ with $\omega' = 1198{\cdot}1$ cm^{-1} and $\omega'' = 1037{\cdot}2$ cm^{-1} (Pearse and Gaydon 1950). Recently Ogawa (1954) has reported a somewhat similar band system in NO excited by silent discharge. Since the preparation of our phosphors was carried out in air it is conceivable that NO could be responsible for the effects. Some experiments carried out since the publication of Ogawa's paper, however, seem to eliminate this possibility. Specpure KCl heated with about 1% by weight of analar $KClO_3$ in a high vacuum gave a phosphor showing the same bands as those obtained in the previously described experiments.

If the attribution of our bands to O_2^+ be accepted, an explanation of the thermoelectric behaviour of these phosphors becomes possible. It will be observed (figure 6) that the thermoelectric effects are anomalous in that above the temperature at which the power changes sign there is a rapid increase with temperature. Crawford and Williams (1950) report a similar increase with temperature for the semiconducting zinc fluoride–manganese phosphors although there is no reversal in sign in their results. Granville and Hogarth (1951) show thermoelectric power curves for semiconductors which are very similar to those obtained for our oxygen activated alkali halide phosphors. Their results were considered to indicate that this behaviour was due to surface effects. In our case if it is assumed that the oxygen is adsorbed on active sites on the freshly formed crystal surfaces, this gas would provide available electron levels at the surface lower than the conduction band of the crystal. The energy levels may be so modified near the surface that the filled band may be raised sufficiently to donate electrons to the oxygen thus creating holes in the filled band. At lower temperatures there should be as many electrons as holes taking part so that there will be only a very small negative thermoelectric power due to the greater mobility of the electrons. As the temperature rises, more and more electrons will pass to the adsorbed oxygen until the oxygen band is more than half full. Then the carriers become effectively all holes and the power becomes positive and increases rapidly with temperature. In allocating the effects to ionized oxygen molecules it is true that it might appear unlikely that the gas could be ionized when, for example, the phosphors are prepared by dropping the molten halide into liquid oxygen. But, as de Boer (1935) and others have pointed out, when a molecule is adsorbed on a highly polar crystal surface such as the alkali halides used here, the potential energy–configuration curves may be so greatly modified that the ionization energy may be quite small. The excitation spectra (figures 1 and 2) show that the energy levels of the adsorbed gas must be indeed considerably modified and differently by the different halide crystal surfaces. Considering the luminescent efficiency–temperature results (figures 3, 4 and 5) it is noted that the high temperature thermal de-activation energies are similar for the chlorides and for the bromides. This suggests that the active sites on which the oxygen may be adsorbed may be anion planes and this may account also for the fact that it is difficult to obtain good phosphors of the same type with the fluorides.

---

† The 'surfaces' are taken to include internal surfaces, cracks and 'sub-structure' boundaries.

The wave numbers of the peaks of the components of the emission bands (see table) suggest a progression derived from transitions from a succession of levels of the lower state to the lowest level of the upper state. In the case of the KCl and NaCl phosphors at 20°K it is noted that the spectrum becomes more complicated (or better resolved) by the appearance of two diffuse components between the sharp components. The sharp components appear in all the spectra of all the phosphors and are clearly the progression indicated above. The diffuse components appear only with KCl and NaCl. It is suggested that they might be accounted for by supposing that either (*a*) the levels of O_2^+ are split into two by a Stark effect of the crystal field or (*b*) lattice vibrations are superposed on the electron vibration spectrum of the adsorbed gas. Further analysis with better resolution is however required to test these suggestions. A diagrammatic representation of these possibilities is shown in figure 7.

The fluorides appear to behave somewhat differently from the chlorides and bromides. For although at 77°K both KF and NaF show the same wave number difference ($\sim 1000\,\mathrm{cm}^{-1}$) between the main members of the band system, at 20°K NaF gives a very complicated emission spectrum. With the resolution available at present it is impossible to make any reliable analysis or detailed interpretation of this spectrum except to give in the table the wave numbers of the most obvious peaks.

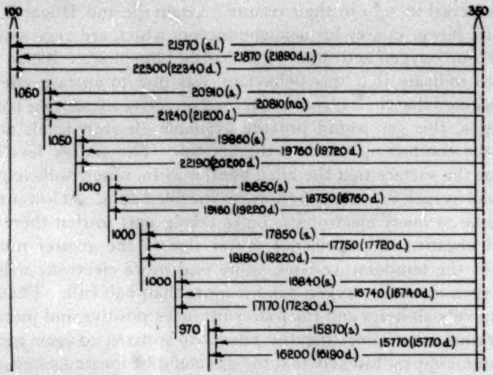

100
350
21970 (s.l.)
21870 (21890d.l.)
22300(22340d.)
1060
20910(s.)
20810(n.o.)
21240(21200d.)
1050
19860(s.)
19760 (19720d.)
22190(20200d.)
1010
18850(s.)
18750(18760d.)
19180 (19220d.)
1000
17850 (s.)
17750 (17720d.)
18180 (18220d.)
1000
16840(s.)
16740 (16740d.)
17170 (17230d.)
970
15870(s.)
15770(15770d.)
16200 (16190d.)

Figure 7. Suggested energy level scheme. (Observed peaks in brackets.)

ACKNOWLEDGMENTS

We are grateful to the Department of Physics of the University of Leeds for facilities afforded and to the Government Grant Committee of the Royal Society for a grant with which some of the apparatus used was purchased. One of us (D. S. B.) is indebted to the Department of Scientific and Industrial Research for a maintenance grant. Professor R. W. B. Pearse was kind enough to discuss some of the spectra with us by letter.

REFERENCES

DE BOER, J. H., 1935, *Electrons and Emission Phenomena* (Cambridge : University Press).
CRAWFORD, J. H., and WILLIAMS, F. E., 1950, *J. Chem. Phys.*, **18**, 775.
EWLES, J., 1949, *Proc. Leeds Phil. Soc.*, **5**, 236.
EWLES, J., and CURRY, C., 1950, *Proc. Phys. Soc.* A, **63**, 708.
EWLES, J., and LEE, N., 1953, *J. Electrochem. Soc.*, **100**, 382.
GRANVILLE, J. W., and HOGARTH, C. A., 1951, *Proc. Phys. Soc.* B, **64**, 488.
VON HIPPEL, A., 1936, *Z. Phys.*, **101**, 680.
HONRATH, W., 1937, *Ann. Phys., Lpz.*, **29**, 421.
MOTT, N. F., 1938, *Proc. Roy. Soc.* , **167**, 384.
OGAWA, M., 1954, *Science of Light* (*Tokyo*), **3**, 39, 54.
PEARSE, R. W. B., and GAYDON, A. G., 1950, *The Identification of Molecular Spectra* (London: Chapman and Hall).

RESEARCH NOTES

Thermal Expansion of Caesium Bromide

By R. S. KRISHNAN and R. SRINIVASAN

Department of Physics, Indian Institute of Science, Bangalore, India

Communicated by R. Fürth; *MS. received* 8*th December* 1955 *and in revised form* 24*th January* 1956

§ 1. Introduction

Caesium bromide has come to occupy an increasingly important position in optical instrumentation because of the wide range of infra-red transmission (up to 40 μ) that it possesses. Ballard, Combes and McCarthy (1952) have made a comparison of the physical properties of this material with those of KRS5 and have mentioned the advantages of caesium bromide over the other crystal.

Caesium bromide has the body-centred cubic ionic structure. A study of thermal expansion of this crystal would therefore also be of theoretical interest.

Menary, Ubbelohde and Woodward (1951) have studied the thermal transition in caesium chloride. They have measured the lattice constant of caesium chloride as a function of temperature from 20 to 469°c. Johnson, Agron and Bredig (1955) have measured the lattice constants of caesium bromide and iodide as a function of temperature up to its melting point. But the values of the expansion coefficient at different temperatures deduced by them are not of great accuracy. In the case of caesium iodide, their value of the linear expansion coefficient at 25°c is 20% smaller than the value of Rymer and Hambling (1951) determined with a precision x-ray method.

In view of its importance in optical instrumentation it was thought desirable to measure the expansion coefficient of caesium bromide with the interference method capable of yielding great precision.

§ 2. Method of Measurement

The interference method suggested by Peters (1919) has been used for measuring the expansion of this crystal. In this method three pieces were cut in the required direction from the crystal and they were ground and polished to form three small pyramids of equal height. Two fused quartz plates with optically polished faces formed the interferometer plates, the three pyramids of the crystal acting as spacers between them. The most stable of the pyramids was also loaded with a U-shaped weight to prevent it from tilting, as suggested by Saunders (1929). When a parallel beam of monochromatic light was allowed to fall on this system straight line interference fringes of width determined by the slight difference in height between the specimens were formed. They could be observed through a suitable optical arrangement. As the temperature was raised the expansion of the specimens would cause the fringes to move across a fixed mark. From a knowledge of the number of fringes moving for a known temperature interval, the coefficient of linear expansion of the specimen could be evaluated.

The apparatus used in the present investigation for work above room temperature was that described by Press (1949). The essential component was the furnace which was carefully designed to prevent large temperature gradients along its axis. The interferometric arrangement was housed in the furnace. The junction of a calibrated chromel–alumel thermocouple in contact with the under surface of the bottom interferometer plate measured the temperature of the crystal. The furnace was heated gradually, the rate of rise of temperature being 0·75°c per minute and the thermo-e.m.f. of the thermocouple was measured every minute on a precision vernier potentiometer reading up to 1 μv. This apparatus has been used for measuring the expansion coefficient of a number of substances by Press (1949), Sharma (1950 a, b, 1951 a, b) and Srinivasan (1955) and very good agreement with the values in the literature have been observed showing that there was insignificant lag in temperature between the crystal and its environment for this rate of heating. The rate of heating was controlled and made uniform. The movement of the fringes was observed visually and the time of transit of the fringes across a reference mark was noted to the nearest second. The temperature at the time of transit of the fringe could be found by interpolation. Before an experiment was begun, the interferometer was heated to the highest temperature to which the experiments were intended to be performed and cooled back gradually. In these measurements, a small correction for the variation of the refractive index of the air with temperature was applied.

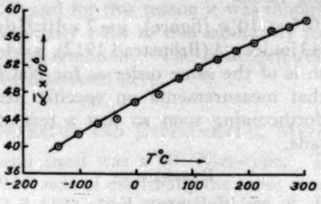
The apparatus used for low temperature measurements was that described by Sreedhar (1952). The interferometer was housed in a central hole bored in a brass block weighing 4 kg. A small heater was enclosed in another hole in the block. A calibrated copper–constantan thermocouple was used for measuring the temperature. The interferometer chamber was evacuated to about 4 mm of mercury to prevent the deposition of solid carbon dioxide and ice on the interferometer plates, at the same time providing good thermal contact between the thermocouple and the interferometer plates. No correction for the variation of refractive index of air with temperature was applied for low temperature measurements. A copper vacuum jacket surrounded the brass block. The interferometer was first cooled to liquid air. After the block had reached the equilibrium temperature, the outer jacket was evacuated. The current in the heater was switched on and observations of the temperature and fringe shift were made as the temperature of the brass block rose gradually till it attained room temperature.

With this apparatus Srinivasan (1955) has determined the thermal expansion coefficients of sodium chloride and iron pyrites. For these two crystals measurements have already been made in this temperature range by Buffington and Latimer and Valentiner and Wallott respectively. The agreement between the values determined with this apparatus and the values for these two crystals in the literature is excellent and shows conclusively that there is insignificant lag in temperature between the crystal and its environment with the rate of heating used.

The single crystal of caesium bromide employed in this investigation was of synthetic origin and grown from melt. The specimen was procured from Messrs. Hilger and Watts Ltd., through the kind courtesy of Dr. A. C. Menzies, Research Director, to whom the author's thanks are due. Three pyramids of 0·367 cm height were prepared from this crystal and were used as the spacers in the interferometer.

§ 3. RESULTS

The mean values of the linear expansion coefficient $\bar{\alpha}$, at various temperatures obtained from these measurements, are collected in table 1. The variation of $\bar{\alpha}$ with temperature is illustrated graphically in the figure. The measurements are accurate to about 3%. The values of the coefficient of expansion are mean values over temperature intervals of 16 centigrade degrees.



($\bar{\alpha}$, T) curve for CsBr.

Table 1. Mean Linear Expansion Coefficient $\bar{\alpha}$ of CsBr

T (°c)	$\bar{\alpha}\times 10^6$	T (°c)	$\bar{\alpha}\times 10^6$
−139·3	40·0	110·9	51·7
−102·9	41·7	144·5	52·9
−67·9	43·3	177·3	54·1
−33·9	44·0	209·2	55·2
−1·3	46·6	240·4	57·0
40·0	47·7	270·8	57·7
76·2	50·1	300·8	58·7

The coefficient of expansion for CsBr can be well represented over the entire range, i.e. from − 139° to 300°c by the following equation:

$$\alpha_t = (46{\cdot}60 + 4{\cdot}67\times 10^{-2}t - 1{\cdot}78\times 10^{-5}t^2)\times 10^{-6}.$$

where t is the temperature in degrees c.

Baxter and Wallace (1916) have measured the volume expansion coefficient of CsBr employing the pyknometer-technique. The mean linear expansion coefficient between 0 to 50°c was found by them to be 47×10^{-6} which agrees well with the value $47{\cdot}7\times 10^{-6}$ at 25°c read from the figure.

The expansion coefficient values determined by us is compared with the values determined by Johnson, Agron and Bredig (1955) with an x-ray method.

Table 2

T (°c)		25	179	208	270	297
$\bar{\alpha}=10^6$	present values	47·6	54·1	55·2	57·7	58·7
	Johnson *et al.*	39·0	49·8	51·8	56·0	57·9

We find that the present values are consistently higher than the values reported by these authors and the difference amounts to about 20% at 25°c. We must mention that in CsI at 25°c these authors obtained a value of $40{\cdot}4\times 10^{-6}$ for α; the value of Rymer and Hambling (1951) determined with a precision x-ray method is $48{\cdot}6\times 10^{-6}$ and the difference amounts to about 20%. Yet these authors claim that the agreement is good.

The Grüneisen constant γ is calculated from the equation

$$\gamma = \beta / \rho \chi c_v$$

where β is the volume expansion coefficient of the crystal, χ its compressibility, c_v the specific heat per g of the substance and ρ its density.

Taking the following values :

$\beta = 3\alpha = 141 \times 10^{-6}$ at 10°c (figure), $\chi = 7 \times 10^{-12}\,\mathrm{dyn\,cm^{-2}}$ (Richards and Saerens 1924), $c_v = 0{\cdot}243\,\mathrm{joules\,g^{-1}}$ (Brönstead 1912), $\rho = 4{\cdot}44\,\mathrm{g\,cm^{-3}}$, γ comes out as equal to 1·87, which is of the same order as for NaCl and KCl (Srinivasan 1955). It is hoped that measurements on specific heat and other physical properties would be forthcoming soon so that a test of Grüneisen's law for this crystal could be made.

REFERENCES

BALLARD, S. S., COMBES, L. S., and MCCARTHY, K. A., 1952, *J. Opt. Soc. Amer.*, **42**, 65.

BAXTER, G. P., and WALLACE, C. C., 1916, *J. Amer. Chem. Soc.*, **38**, 259.

BRÖNSTEAD, 1912, *Z. Elektrochem.*, **18**, 714.

JOHNSON, J. W., AGRON, P. A., and BREDIG, M. A., 1955, *J. Amer. Chem. Soc.*, **77**, 2734.

MENARY, J. W., UBBELOHDE, A. R., and WOODWARD, I., 1951, *Proc. Roy. Soc.* A, **208**, 158.

PETERS, C. G., 1919, *J. Wash. Acad. Sci.*, **9**, 281.

PRESS, D. C., 1949, *Proc. Ind. Acad. Sci.* A, **30**, 284.

RICHARDS, W., and SAERENS, E. P. R., 1924, *J. Amer. Chem. Soc.*, **46**, 934.

RYMER, T. B., and HAMBLING, P. G., 1951, *Acta Cryst., Camb.*, **4**, 565.

SAUNDERS, J. B., 1929, *J. Res. Nat. Bur. Stand.*, **23**, 179.

SHARMA, S. S., 1950 a, *Proc. Ind. Acad. Sci.* A, **31**, 83, 261, 339 ; 1950 b, *Ibid.*, **32**, 268, 285 ; 1951 a, *Ibid.*, **33**, 245, 283; 1951 b, *Ibid.*, **34**, 72.

SREEDHAR, A. K., 1952, *Proc. Ind. Acad. Sci.* A, **36**, 141.

SRINIVASAN, R., 1955, *J. Indian Inst. Sci.*, **37**, 232.

Electron Impact Studies of the Dissociation Energies of N_2 and H_2

BY R. THORBURN AND J. D. CRAGGS

Department of Electrical Engineering, University of Liverpool

MS. received 24th January 1956

§ 1. INTRODUCTION

Of the various processes which can take place when a molecule XY is subjected to electron bombardment that which has been studied in this work and which is the most important in the evaluation of the dissociation energy of the molecule XY, i.e. $D(\mathrm{XY})$ is the dissociative ionization process

$$\mathrm{XY} + \mathrm{e} \rightarrow \mathrm{X}^+ + \mathrm{Y} + 2\mathrm{e}. \qquad \ldots\ldots(1)$$

Denoting the appearance potential of X^+, the ionization potential of X and the dissociation energy of XY by $A(\mathrm{X}^+)$, $I(\mathrm{X})$ and $D(\mathrm{XY})$ respectively, the relationship between these quantities for the process (1) is

$$A(\mathrm{X}^+) = D(\mathrm{XY}) + I(\mathrm{X}) + K + E \qquad \ldots\ldots(2)$$

where $K+E$ represents the sum of the kinetic and excitation energies possessed by the fragments involved in these processes. From equation (2) it is seen that if the values of $A(\mathrm{X}^+)$, $I(\mathrm{X})$ and $K+E$ are known then the value of $D(\mathrm{XY})$ can be determined.

The present work was prompted by the existence of widely differing values for $D(N_2)$. Gaydon (1947) summarized the position for both the spectroscopic and electron impact data, discussed the three most likely values, i.e. 9·764 ev, 8·573 ev, and 7·383 ev and stated that the first seemed the most likely. The electron impact work by itself has given a similar range of values (see, for example, Massey and Burhop 1952) and for this reason it was thought worth while to use the Lozier apparatus (originally built in this laboratory for different purposes) in an attempt to clarify the situation. The hydrogen results were taken as a check and are only briefly discussed here.

§ 2. Apparatus and Experimental Methods

The mass spectrometer used was a 60° Nier-type. The Lozier apparatus (Marriott and Craggs 1954) was of essentially the same construction as that used by Lozier (1934) and need not be described here.

Two types of measurement have been carried out in this work. Firstly, the positive ion current as function of electron energy and secondly the positive ion current as a function of the radial retarding potential in the Lozier apparatus have been determined. These two measurements yield the appearance potentials of the positive atomic ions and their kinetic energies. Since the Lozier apparatus does not give mass analysis the appearance potential of the atomic ion is determined by observing a 'break' in the ionization efficiency curve. The difference in electron energy between this 'break' and the onset of the parent ion formation is equivalent to the difference in appearance potentials for the diatomic and atomic ions. The standards used for energy scale calibration are the ionization potentials of the respective molecules, i.e. of N_2 and H_2. Due to the shape of the ionization efficiency curves for these molecules and because of the relatively small amount of atomic ion formation the only method of deducing the appearance potentials of the atomic ions is that of extrapolated linear intercepts (see Barnard 1953). The validity of this method has been checked in nitrogen with results which are given below. The data given for nitrogen and hydrogen have been obtained using the Lozier apparatus, the mass spectrometer being used solely for the identification of a particular ion occurring at some 'break' on the ionization curves obtained with the Lozier apparatus and the procedure outlined above for obtaining these data is justified by the equality of the two estimates of $A(N^+)$ from nitrogen, quoted below.

§ 3. Results and Discussion

Nitrogen.

In nitrogen there are breaks in the ionization efficiency curve, one of them corresponding to the formation of N_2^+ ions in the reaction

$$N_2 + e \rightarrow N_2^+ + 2e.$$

The appearance potential of the N_2^+ ions corresponds to the ionization potential of nitrogen, a value of 15·7 ev. This value has been determined from the mass spectrometer using $A(A^+)$ from argon as the standard (15·76 ev) and is in agreement with previous values given in the literature. The appearance potential for the process giving N^+ in the Lozier apparatus was obtained from this value by the method of extrapolated linear intercepts, and so $A(N^+)$ was found to be $24{\cdot}3 \pm 0{\cdot}3$ ev.

As has been explained above the appearance potential of N^+ from N_2 has also been obtained using the mass spectrometer as a check on the Lozier reading and the value thus found was $24{\cdot}4 \pm 0{\cdot}1$ ev agreeing so well with the Lozier apparatus value of $24{\cdot}3 \pm 0{\cdot}3$ ev that it would seem certain that the 'break' in the ionization curve in the latter apparatus does indeed denote the onset of N^+ formation by the dissociative ionization process

$$N_2 + e \rightarrow N^+ + N + 2e.$$

The results of the application of retarding potentials to the postivie ions formed by 35 volt electrons using argon for the purpose of comparison and for calibrating the retarding potential scale show that within experimental error, which is $\pm 0{\cdot}1$ ev, ions of zero kinetic energy are formed. Thus the appearance potential of the N^+ ions at 24·3 ev corresponds to that of ions of zero kinetic energy. This value is in agreement with those obtained by Lozier (1934) ($24{\cdot}27 \pm 0{\cdot}2$ ev), Hagstrum (1951) ($24{\cdot}3 \pm 0{\cdot}2$ ev) and Vaughan (1931) (24·5 ev).

From equation (2) above we have

$$A(N) = D(N_2) + I(N) + K + E$$

and putting $K = 0$

$$D(N_2) = 9{\cdot}8\,\text{ev} - E.$$

The method of electron impact does not give a unique value of $D(N_2)$ as the degree of excitation of the products of dissociation still remains undetermined. A considerable number of values of $D(N_2)$ have been favoured at various times, the two receiving most support being the 9·8 ev and the 7·4 ev values. The latter were deduced from the electron impact data by assuming that the neutral fragment involved in the dissociation process was in an excited state in the reaction $N_2(x^1\Sigma_g^+) \rightarrow N^+(^3P) + N^*(^2D^0)$. Recent work by Clarke (1954) has shown that the products are unexcited. This would appear to be the most natural interpretation of the electron impact data and fixes the value of $D(N_2)$ at $9{\cdot}8 \pm 0{\cdot}2$ ev.

Fox and Hickam (1954) have studied the formation of N_2^+ at energies just above the critical (appearance) potential which they found to be $15{\cdot}60 \pm 0{\cdot}01$ ev, in good agreement with the spectroscopic value of 15·58 ev. Using a velocity selector, Clarke (1954) studied the formation of N^+ and N_2^+ and found the appearance potential of N^+ to be $24{\cdot}35 \pm 0{\cdot}1$ (xenon as standard) and taking $I(N)$ as 14·54, $D(N_2)$ was found to be 9·79 ev. The spectroscopic value of 9·764 is given by Douglas (1952) who has also summarized the recent values (1955).

Hydrogen.

The dissociation processes produced in hydrogen have already been completely studied by Lozier (1930) and have never been seriously questioned, so they were repeated to provide a check on the performance of the instrument and on the method for obtaining dissociation energies. In the case of hydrogen a correction has to be made for the effect of the magnetic field on the energy of the ions (Lozier 1930).

The ion current curve for ions possessing a minimum kinetic energy of 1·4 ev gave the appearance potential for these ions, determined by the method of extrapolated linear intercepts from the onset of H_2 formation as $20{\cdot}7 \pm 0{\cdot}3$ ev. The total amount of kinetic energy involved is 2·8 ev ($H + H^+$) and from $A(H^+) = D(H_2) + I(H) + 2{\cdot}8$ ev, $D(H_2) = 4{\cdot}4 \pm 0{\cdot}3$ ev.

It is assumed that the products H^+ are unexcited. The value quoted above is in favourable agreement with that found by other workers (see Massey and Burhop (1952) where a value of 4·4 ev is quoted).

References

Barnard, G. P., 1953, *Modern Mass Spectrometry* (London : Institute of Physics)
Clarke, E. M., 1954, *Canad. J. Phys.*, **32,** 764.
Douglas, A. E., 1952, *Canad. J. Phys.*, **30,** 302; 1955, *J. Phys. Chem.*, **59,** 109.
Fox, R. E., and Hickam, W. M., 1954, *J. Chem. Phys.*, **22,** 2057.
Gaydon, A. G., 1947, *Dissociation Energies* (London : Chapman and Hall).
Hagstrum, H. D., 1951, *Rev. Mod. Phys.*, **23,** 185.
Lozier, W. W., 1930, *Phys. Rev.*, **52,** 843 ; 1934, *Phys. Rev.*, **46,** 268.
Marriott, J., and Craggs, J. D., 1954, *E.R.A. Report* L/T308.
Massey, H. S. W., and Burhop, E. H. S., 1952, *Electronic and Ionic Impact Phenomena* (Oxford : University Press).
Vaughan, A. L., 1931, *Phys. Rev.*, **38,** 1687.

A Note on the Electron Distribution Function in Insulators at High Field Strength

By J. J. O'DWYER
Division of Electrotechnology, National Standards Laboratory, C.S.I.R.O., Sydney

MS. received 2nd February 1956

The distribution function of the conduction electrons in an insulator for applied fields of near breakdown strength has been calculated by Fröhlich (1947), Heller (1951) and Franz (1952). In a previous publication (1954) the present author has noted the equivalence of the solutions of Fröhlich and Heller for high energies (but lower than the ionization energy). It was also stated erroneously that Franz had assumed that the effect of the field was to shift the electron distribution bodily in momentum space by an amount determined by some relaxation time constant. The author was led to this belief by Franz's incomplete treatment of the effect of collisions, and it is the purpose of this note to indicate the consideration which has been omitted by Franz and which, if included, reduces his solution also to the Fröhlich solution.

In considering the effect of electron lattice collisions Franz (1952, eqn (1)) gives for the current of electrons in energy space caused by these collisions

$$S(E)\,|_{\mathrm{st}}=\sum\int_{E-\epsilon}^{E}\chi w_{\mathrm{a}}\,dE-\sum\int_{E}^{E+\epsilon}\chi w_{\mathrm{e}}\,dE \qquad \ldots\ldots(1)$$

where $\chi(E)\,dE$ is the number of conduction electrons whose energies lie between E and $E+dE$, w_{a} and w_{e} are the probabilities per unit time for the absorption and emission of a lattice quantum respectively, and ϵ is the energy of the lattice quantum. It is stated that the summation Σ is over all collision processes, and that ϵ is relatively small (presumably compared with $\boldsymbol{k}T$) so that (1) is developed as

$$S(E)\,|_{\mathrm{st}}=-\sum\epsilon(w_{\mathrm{e}}-w_{\mathrm{a}})\,\chi(E)-\sum\tfrac{1}{2}\epsilon^{2}\,w_{\mathrm{s}}\frac{d\chi}{dE}(E) \qquad \ldots\ldots(2)$$

where $w_{\mathrm{s}}=w_{\mathrm{e}}+w_{\mathrm{a}}$.

However, Franz seems to have overlooked the fact that the summation itself depends on the energy (e.g., $w_e + w_a$ is not a function of E but $\Sigma\epsilon(w_e + w_a)$ is) so that it appears that the Σ also should be under the integral sign in (1). The development for $\epsilon \ll \boldsymbol{k}T$ would then be

$$S(E)|_{\text{st}} = -\sum \epsilon(w_e - w_a)\,\chi(E) - \sum \tfrac{1}{2}\epsilon^2 w_s \frac{d\chi(E)}{dE} - \chi(E)\frac{d}{dE}\Big[\sum \tfrac{1}{2}\epsilon^2 w_s\Big]. \quad .(3)$$

Since this expression is rather formal it is of interest to evaluate the summations involved. This has been done by Fröhlich (1937) who finds (cf. eqn (17) of that paper)

$$B(E) = \sum \epsilon(w_e - w_a) = \frac{\sqrt{2}\,\pi\,\boldsymbol{e}^4\,\boldsymbol{m}^{1/2}}{Ma^3 E^{1/2}}\log\gamma \qquad \ldots\ldots(4)$$

and using a similar argument we find

$$C(E) = \sum \tfrac{1}{2}\epsilon^2 w_s = \tfrac{1}{2}\epsilon B(E)\left\{1 + \frac{2}{\exp(\epsilon/\boldsymbol{k}T) - 1}\right\} \simeq \boldsymbol{k}T\,B(E). \qquad \ldots\ldots(5)$$

Thus also from (4) and (5)

$$\frac{dC(E)}{dE} \simeq -\frac{\boldsymbol{k}T}{2E}B(E) \qquad \ldots\ldots(6)$$

so that using (4), (5) and (6) in (3)

$$S(E)|_{\text{st}} = B(E)\left[-\chi(E) - \boldsymbol{k}T\frac{d\chi E}{dE} + \frac{\boldsymbol{k}T}{2E}\chi(E)\right]. \qquad \ldots\ldots(7)$$

Fröhlich, using a dimensionless variable $x = E/\boldsymbol{k}T$ and a density function $g(x)$, obtains (1947, eqn (16)) for the current

$$j(x,t)|_{\text{coll}} = \frac{B(x)}{\boldsymbol{k}T}\left[\frac{dg(x)}{dx} + \left(1 - \frac{1}{2x}\right)g(x)\right]$$

which is identical with (7) since $g(x) = \boldsymbol{k}T\,\chi(E)$ and $dg(x)/dx = (\boldsymbol{k}T)^2\,d\chi(E)/dE$. The difference in sign simply means that S and j have opposite directions of positive current flow.

It cannot be objected that this extra term is unimportant since Franz explicitly claims to deal with low energies and it is just this case in which the term is most important since then $\boldsymbol{k}T/2E \gtrsim 1$. Neither can it be objected that the approximations used by Fröhlich in deriving (4) and extended here to (5) and (6) do not deal adequately with the case as proposed by Franz. Franz himself implies $\boldsymbol{k}T \gg \epsilon$ as already seen, and Callen (1949) (whose results Franz quotes to evaluate his summations) gives $E > \epsilon$ (eqn (59) of his paper) and infers for practical calculations on the alkali halides that $E \gg \epsilon$ (discussion of the validity of the perturbation treatment in his paper).

Finally it should be mentioned that ignoring the energy dependence of the 'sum over collision processes' amounts in a certain way to treating the collision relaxation time as constant since it is essentially from this 'sum over collision processes' that it too derives its energy dependence.

REFERENCES

CALLEN, H. B., 1949, *Phys. Rev.*, **76,** 1394.
FRANZ, W., 1952, *Z. Phys.*, **132,** 285 ; 1953, *Ergebn. exakt. Naturw.*, **27,** 1.
FRÖHLICH, H., 1937, *Proc. Roy. Soc.* A, **160,** 230 ; 1947, *Ibid.*, **188,** 532.
HELLER, W., 1951, *Phys. Rev.*, **84,** 1130.
O'DWYER, J. J., 1954, *Aust. J. Phys.*, **7,** 36.

Epitaxial Growth of Crystals of Hexagonal Selenium

By R. A. HYMAN

Standard Telecommunication Laboratories Ltd., Enfield, Middx.

Communicated by R. W. Ditchburn; MS. received 6*th December* 1955, *and in revised form* 21*st February* 1956

Crystals of hexagonal selenium in the form of a cactus were first obtained by Brown (1914): a thin needle, with the *c*-axis parallel to its length forms a spine from which grow an array of branching needles, in each of which the *c*-axis is again parallel to the length; the area of contact is in a prism plane of both the spine and branch needles. The author has grown some crystals of selenium by deposition from the vapour, including a number with this form. The angles between the *c*-axes have been measured and found to correspond to preferred orientations of epitaxial growth.

40 grammes of selenium with a purity of 99·99% were kept molten at 240°c at the bottom of a sealed Pyrex tube, in an atmosphere of spectroscopically pure argon at a pressure of 5 cm Hg. At the top of the furnace, in a region where the temperature had been maintained at 205 ± 10°c, an array of crystals had formed.

Among the crystals were a number of the type shown in figure 1 (Plate). These fall into two groups: in one group (figure 1 (*a*)), of which twenty were found, the *c*-axes of the branching needles were perpendicular to the *c*-axis in the spine; in the other group (figure 1 (*b*)), of which ten examples were measured, the *c*-axes in the branch needles were at an average angle of 40 ± 3° to the *c*-axis of the spine. If a layer of atoms is deposited on a crystal surface of selenium the state of lowest energy clearly occurs when the layer continues the lattice. The next most favourable orientations, shown in figure 2, are when the *c*-axes are at 90° and 41° to the original directions, the misfits being 14% and 25% respectively, both of which lie within the limits of misfit for permitted epitaxial growth. The possibility of epitaxial growth is, in this case, determined solely by the geometry, because the interchain forces are residual and therefore isotropic.

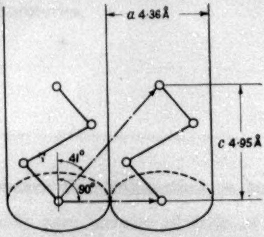


Figure 2. The crystal structure of hexagonal selenium. Chain-dotted lines show preferred directions of *c*-axis in epitaxial growth.

The crystals show cleavage only along prism planes (de Boer 1947) and they are prone to permanent distortion by bending, twisting or simply crushing. The spine of the crystal shown in figure 1 (*b*) was originally straight but was bent before being photographed.

The prevalence of the needle shape in the crystals proves that the rate of growth tends to be greater along the *c*-axis. This is probably the result of the absence of fundamentally flat planes in the selenium structure, other than the prism planes. The absence of a fundamentally flat plane of course removes the restriction of growth to regions including an emergent screw dislocation.

The ring molecules, which are the main constituent of selenium vapour at about 240°C (Preuner and Brockmöller 1912), are not a constituent of hexagonal selenium, and it is necessary for the rings to break as they are added to the hexagonal structure. When there is no free chain-end—the chain might, for example, be terminated by a monovalent impurity atom such as iodine—the probability of the selenium ring-of-eight molecule breaking and being joined to the lattice is reduced by a factor of the order $\exp(E/\boldsymbol{k}T)$, where E is the energy of the Se–Se bond, 2·5 ev, and T is the absolute temperature of the crystal, say 480°K: that is, the probability is reduced by approximately $\exp(50)$. Therefore, it may be concluded that in the case of selenium the crystals only grow rapidly where there is a free chain end.

ACKNOWLEDGMENTS

Most of this work was carried out during leave of absence at Reading University and the author would like to thank Professor Ditchburn for the provision of facilities to carry out this work, and Dr. H. K. Robinson for many useful discussions.

REFERENCES

DE BOER, F., 1947, *Philips Res. Rep.*, **2**, 349.
BROWN, F. C., 1914, *Phys. Rev.*, **4**, 85.
PREUNER, G., and BROCKMÖLLER, I., 1912, *Z. Phys. Chem.*, **81**, 129.

LETTERS TO THE EDITOR

A Simple Etching Technique for Revealing Dislocations in Silicon

A method has been developed for revealing dislocations in crystals of silicon by chemical etching.

CP–4, the etchant used for this purpose on germanium (Vogel *et al.* 1953, Oberly 1954), did not in fact produce dislocation etch pits on silicon and accordingly a series of experiments was made until a suitable etching technique was obtained for this purpose.

The following procedure was found to give clear reproducible results as illustrated in the figure (Plate), which shows an etch pit pattern on a (111) crystal face. Pits were also obtained on planes within 10° of the (110) and (100) faces.

The crystal face to be examined was ground flat on a glass plate using 302 grade carborundum powder and then etched in a bath containing HF(40%) 30–40%, HNO_3(conc.) 30–35%, CH_3COOH (glacial) 30–35%, until the surface was polished. This took 1–2 minutes. The etchant was then diluted with 40–60% water, whilst the crystal remained in the bath. After $\frac{1}{2}$–$1\frac{1}{2}$ minutes the crystal was removed and on examination the etch pits were visible. The sizes of the etch pits were to some extent dependent on the time of etching and the dilution; if the dilution was too great then the crystal surface tarnished obliterating the etch pits formed.

The etch pit densities counted on the crystals varied between 10^4 and 10^5 cm^{-2}. These are comparable with values given by Kurtz, Kulin and Averbach (1956) for edge dislocations in germanium and silicon.

Kurtz, A. D., Kulin, S. A., and Averbach, B. L., 1956, *Phys. Rev.*, **101,** 1285.
Oberly, J. J., 1954, *J. Metals, N.Y.*, **6,** 1025.
Vogel, F. L., Pfann, W. G., Corey, H. E., and Thomas, E. E., 1953, *Phys. Rev.*, **90,** 489.

Mullard Research Laboratories,
Salfords, Redhill,
Surrey.
23rd April 1956.

C. A. F. T. Spray.

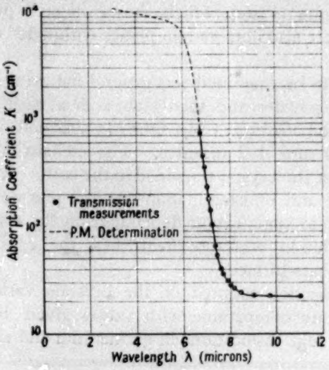
Radiative Recombination in InSb

The application to fairly pure samples of InSb of the theory of radiative recombination of holes and electrons in a semiconductor, as given by van Roosbroeck and Shockley (1954) requires an accurate knowledge of the absorption spectrum for wavelengths greater than about 3 μ. Previous calculations on radiative recombination by Avery and Jenkins (1955) and by Mackintosh and Allen (1955) were based on absorption spectra which were incomplete due to the difficulties which arise in carrying out transmission and reflection measurements in the region with absorption coefficient 10^3–10^4 cm^{-1}.

It has been found possible to obtain values of the absorption coefficient in this region by making a study of the short circuit photomagnetic current per unit quantum as a function of wavelength (paper to be published later). The complete absorption spectrum at room temperature has been obtained in this way for a single crystal of InSb containing 4×10^{15} extrinsic holes per cm^3 and is shown in the figure. The optical resolution for transmission measurements was $0{\cdot}07\,\mu$. The radiative recombination rate R calculated from this spectrum is

$$R = 2{\cdot}25 \times 10^{22}\ \mathrm{cm}^{-3}\ \mathrm{sec}^{-1}$$

at room temperature and due to the purity of the crystal this can be taken as the recombination rate in intrinsic material.



Absorption spectrum of InSb at room temperature

Taking the electron or hole concentration in intrinsic InSb at room temperature to be $n_i = 1{\cdot}6 \times 10^{16}\ \mathrm{cm}^{-3}$, the value of the radiative recombination lifetime of carriers injected into intrinsic material is

$$\tau_i = n_i/2R = 0{\cdot}36\ \mu\mathrm{sec}.$$

This value lies somewhat lower than those given by previous workers and is more accurately determined, the total error being not more than 10%. Errors arise from a lack of knowledge of an accurate value of n_i and because of the finite resolution of the optical system.

Acknowledgment is made to the Controller, H.M. Stationery Office, for permission to publish this letter.

Radar Research Establishment, D. W. Goodwin.
Great Malvern, T. P. McLean.
Worcs.
10th February 1956.

Avery, D. G., and Jenkins, D. P., 1955, *J. Electronics*, **1,** 145.
Mackintosh, I. M., and Allen, J. W., 1955, *Proc. Phys. Soc.* B, **68,** 985.
van Roosbroeck, W., and Shockley, W., 1954, *Phys. Rev.*, **94,** 1558.

The Magnetic Properties of MnAs and MnBi

It is well known that the intermetallic compounds MnAs and MnBi which crystallize with the NiAs hexagonal structure are ferromagnetic at room temperature and that they undergo transitions, at 45°c and 360°c respectively, to magnetic states which show some of the characteristics of antiferromagnetism (Guillaud 1951).

Associated with the magnetic changes are discontinuous changes in lattice constants (Willis and Rooksby 1954) and for MnAs a specific heat anomaly of the λ type has been observed near 45°c (Bates 1927, Guillaud 1951).

Street and Bacon (1955) have made a neutron diffraction study of MnAs and have failed to observe superstructure lines or changes in intensity in the (0002) and $(10\bar{1}1)$ reflections due to antiferromagnetism above 45°c in a polycrystalline specimen. Their results indicate that in the ferromagnetic state the moments of the manganese atoms are orientated in a direction perpendicular to the *c* axis. They have studied the $(10\bar{1}0)$ reflection in a single micro-crystal and observed a decrease of about a quarter in the intensity of this reflection on raising the temperature through 45°c. The absence of evidence of antiferromagnetic ordering is consistent with other evidence about the properties of these compounds, which suggests that the magnetic interactions are always ferromagnetic, and a recent study (Leech and Manuel 1956) of the properties of some anhydrous chlorides or the iron group has indicated how a two-dimensional ferromagnetic ordering can simulate antiferromagnetism because of the high shape anisotropy of single layer domains.

MnAs is paramagnetic above 126°c and follows a Curie–Weiss law with a paramagnetic Curie temperature of 283°K (10°c) (Serres 1947). This high positive value of θ indicates a very large internal field favouring ferromagnetic ordering in the temperature region above 126°c. There is no discontinuous structural change at 126°c and below this temperature a ferromagnetic and not an antiferromagnetic internal field should exist. Heikes (1955) has shown similarly that the internal field in MnBi above 633°K is of a ferromagnetic nature.

The lattice of manganese atoms in these compounds is made up of superimposed layers of triangular plane lattices in which at room temperature immediate neighbours are at a distance of about 3·72 Å in MnAs and 4·28 Å in MnBi. The distances between immediate neighbours in adjacent layers of manganese atoms are about 2·85 Å in MnAs and 3·05 Å in MnBi. There is the possibility that the large positive values of θ in the Curie–Weiss Law are consistent with a strong ferromagnetic interaction in one direction accompanied by a weak antiferromagnetic interaction in a perpendicular direction. The symmetry of the lattice does not favour antiferromagnetic ordering in basal planes (Wannier 1950) and the only plausible antiferromagnetic structure would be one in which the basal planes were ordered ferromagnetically, adjacent basal planes having their moments directed in anti-parallel directions. The large separation of immediate neighbours in the basal planes (3·72 Å, 4·28 Å) is consistent with a ferromagnetic exchange interaction between the d shells of the manganese atoms.

If there is to be a ferromagnetic–antiferromagnetic transformation at 45°c as postulated by Guillaud (1951) there should be a change in sign of the exchange interaction at this temperature due to change in the separation of the magnetic atoms.

As the temperature of MnAs is increased through 45°c there is a discontinuous change in a which decreases to about 3·68 Å, still very large and favouring ferromagnetic ordering, and a steady expansion in the direction of the c axis, the increasing separation of the magnetic atoms being inconsistent with a change from ferromagnetic to antiferromagnetic interaction. For MnBi there are discontinuous changes at the transition temperature, a increasing and c decreasing. The separation of the manganese atoms in the c direction decreases to 2·99 Å which is greater than the separation of the manganese atoms in ferromagnetic MnAs.

Since there appears to be ferromagnetic ordering on both sides of the transition temperatures in the compounds in question, a change from one to another kind of ferromagnetic ordering is indicated. It is suggested that at low temperatures the substances are ferromagnetically ordered in three dimensions, but on raising the temperature a relatively weaker interaction between the basal planes is broken up by thermal agitation leaving an assembly of two-dimensional ferromagnetically ordered triangular networks of the basal planes; the direction of the magnetic moment of each basal plane would vary randomly from plane to plane but would always remain normal to the c axis, on account of shape anisotropy of the single layer ferromagnetic domains. Such a magnetic structure could give rise to magnetic properties similar to those of antiferromagnetic substances (Leech and Manuel 1956).

Kurti (1951) has observed that the entropy of ordering is of the same magnitude for both ferromagnetism and antiferromagnetism and a simple change from one to the other should not involve a specefic heat anomaly of the size measured. A change from three-dimensional to two-dimensional ordering would introduce an entropy change and a corresponding specific heat anomaly, although the discontinuous changes in lattice constants in MnAs at the transition temperature makes the specific heat anomaly less significant in the interpretation of the magnetic transition.

The observed decrease in intensity of the reflection from the $(10\bar{1}0)$ plane in the neutron diffraction experiments of Street and Bacon on MnAs is consistent with the above interpretation if the directions of easy magnetization below and above the transition temperature bear the same relation to the hexagonal symmetry of the crystal structure.

A. J. Manuel.

Department of Physics,
Bedford College,
London.

19th March 1956.

Bates, L. F., 1927, *Proc. Roy. Soc.* A, **67**, 680.
Guillaud, C., 1951, *J. Phys. Radium*, **12**, 223.
Heikes, R. R., 1955, *Phys. Rev.*, **99**, 446.
Kurti, N., 1951, *J. Phys. Radium*, **12**, 227.
Leech, J. W., and Manuel, A. J., 1956, *Proc. Phys. Soc.* B, **69**, 220.
Serres, A., 1947, *J. Phys. Radium*, **8**, 146.
Street, R., and Bacon, G., 1955, *Nature, Lond.*, **175**, 518.
Wannier, G. H., 1950, *Phys. Rev.*, **79**, 357.
Willis, B. T. M., and Rooksby, H. P., 1954, *Proc. Phys. Soc.* B, **67**, 290.

A New Method for Determining the Second Virial Coefficient of a Gas

The dependence of the refractive index n on the molar volume V of a gas is conveniently expressed in the form

$$\frac{n^2-1}{n^2+2}V=P^0\left(1+\frac{\beta}{V}+\frac{\gamma}{V^2}+\dots\right) \qquad \dots\dots(1)$$

where P^0 is independent of the temperature while β, γ, ... depend on the temperature. The form of (1) is analogous to the standard expression for the pressure p

$$pV=\boldsymbol{R}T\left(1+\frac{B}{V}+\frac{C}{V^2}+\dots\right) \qquad \dots\dots(2)$$

where the virial coefficients B, C, ... depend on the temperature. From the analogy between (1) and (2) the coefficients β, γ, ... might also reasonably be called refractivity virial coefficients, but we deplore the recent suggestion by Buckingham and Pople (1955) to attach the name virial coefficients to $P^0\beta$, $P^0\gamma$, ... instead of to β, γ, ...

We here confine our discussion to densities so low that formulae (1) and (2) may with sufficient accuracy be replaced by

$$\frac{n^2-1}{n^2+2}V=P^0\left(1+\frac{\beta}{V}\right). \qquad \dots\dots(3)$$

$$pV=\boldsymbol{R}T\left(1+\frac{B}{V}\right). \qquad \dots\dots(4)$$

Dividing (3) by (4) we have to the same degree of accuracy

$$\frac{n^2-1}{n^2+2}\,\frac{\boldsymbol{R}T}{p}=P^0\left(1+\frac{\beta-B}{V}\right). \qquad \dots\dots(5)$$

The theoretical evaluation of β is considerably more difficult than that of B. The problem was first attacked by Kirkwood (1936) who showed that for non-attracting rigid spheres

$$\beta=\frac{P^{02}}{b}\equiv\frac{P^{02}}{2\pi\boldsymbol{L}\sigma^3/3} \qquad \dots\dots(6)$$

where σ is the molecular diameter, $\boldsymbol{L}$ is Avogadro's number and b is related to σ in the manner of van der Waals. The evaluation of β for an inverse 6–12 interaction has been performed by de Boer, van der Maesen and ten Seldam (1953). Their formula was subsequently rederived by Buckingham and Pople (1955). For our present purpose it is sufficient to write this formula as

$$\frac{\beta}{b}=\left(\frac{P^0}{b}\right)^2\phi\left(\frac{\boldsymbol{k}T}{\epsilon}\right) \qquad \dots\dots(7)$$

where ϵ is the depth of the potential energy at its minimum. All we here need to say about the function ϕ is that at ordinary experimental temperatures it is comparable with unity.

A rough estimation of b and P^0 is sufficient to establish that for A, Xe, H_2, N_2, CH_4, C_2H_4, CO_2 the ratio $(P^0/b)^2$ is about 0·01 or less and we presume that this is also true for other non-polar gases, while B is, except near the Boyle point, comparable with b in magnitude. It follows that in the region of densities where formula (4) is adequate, β is less than the experimental inaccuracy in B. Although the statement that β is negligible compared with B seems not to have been made before, it is borne out by the analysis of experimental measurements at higher pressures (20 to 100 atmospheres) by Miller (1948).

Consequently we may without loss of accuracy replace (5) by

$$\frac{n^2-1}{n^2+2}\frac{RT}{p}=P^0\left(1-B\frac{p}{RT}\right). \qquad \ldots\ldots(8)$$

Formula (8) gives us a new method for evaluating the second virial coefficient B from measurements of n and p without any need for density measurements. This procedure has been applied to ethylene. Measurements of refractive index were made with a Rayleigh interferometer and details will be published later. The experimental results can be expressed by formula (8) with $P^0 = 10{\cdot}734\ \mathrm{cm^3\,mole^{-1}}$ as compared with the previously measured value (Watson and Ramaswamy 1936) $10{\cdot}75\ \mathrm{cm^3\,mole^{-1}}$ and values of B which we compare with those deduced from the density/pressure measurements of Michels and Geldermans (1942) at pressures of 20 atmospheres upwards.

Temperature (°c)		26·6	50·2	63·9
B (cm³ mole⁻¹) from	refractive index	-145 ± 4	-124	-107
	density/pressure	-139	-118	-107

Note added in proof. Measurements on neopentane between 25°c and 70°c have now been made. The values obtained lie between -950 and $-700\ \mathrm{cm^3\,mole^{-1}}$ and agree within about $50\ \mathrm{cm^3\,mole^{-1}}$ with those found by density/pressure measurements.

Department of Chemistry,
University of Reading.
6th February, 1956.

H. M. Ashton.
E. A. Guggenheim.

de Boer, J., van der Maesen, F., and ten Seldam, C. A., 1953, *Physica*, **19**, 265.
Buckingham, A. D., and Pople, J. A., 1955, *Trans. Faraday Soc.*, **51**, 1029.
Kirkwood, J. G., 1936, *J. Chem. Phys.*, **4**, 592.
Michels, A., and Geldermans, M., 1942, *Physica*, **9**, 967.
Miller, J. G., 1948, *Trans. Amer. Soc. Mech. Eng.*, **70**, 645.
Watson, H. E., and Ramaswamy, K. L., 1936, *Proc. Roy. Soc.* A, **156**, 144.

REVIEWS OF BOOKS

The Theory of Hydrodynamic Stability, by C. C. LIN. Pp. xi + 155. (Cambridge: University Press, 1955.) 22*s*. 6*d*.

It is well known that the textbook cases of steady motion of a viscous fluid are difficult to reproduce in practice, and that an unsteady turbulent motion often occurs in their place. Closer observation shows that these states of steady laminary flow are unstable, and, as with the rising plume of hot smoke from a cigarette, develop periodic oscillations of increasing amplitude which ultimately break up the steady motion. Study of the conditions under which a steady laminary motion is unstable to suitable disturbances constitutes the difficult subject of hydrodynamic stability. It has attracted mathematicians for many years, but it did not really come into its own until aeronautical engineers discovered that the frictional drag of a body of streamline form depended vitally on whether the viscous boundary layer near the body surface was unstable to small disturbances. For the most part, engineers now have the required information about the stability of fluid flows, and the rest of us have been presented with a most interesting minor branch of mathematical physics.

By writing this book, the first in the field, Professor Lin has done a considerable service to those interested in fluid mechanics. He has certainly earned the right to write it. Beginning in 1945 with a long paper which infused new life into the subject, he has made contributions to almost all aspects of the theory of hydrodynamic stability, and the book is woven round these contributions. His concern is primarily, although by no means exclusively, for the mathematical analysis of stability problems, and the book provides a careful and illuminating account of the many difficulties that have given rise to so much doubt and confusion in the journal literature.

The bulk of the book is devoted to an analysis of two simple flow systems, which provide prototypes of two kinds of instability. The first is flow in the space between two concentric circular cylinders in steady relative motion, which is unstable to axisymmetric disturbances of suitable size when the circulation at the inner cylinder is a little greater than that at the outer cylinder. The second is plane Poiseuille motion—flow between parallel plane boundaries under the action of a uniform pressure gradient—which is unstable for reasons that are not at all obvious. Professor Lin explains clearly how, in this latter case, the action of viscosity, normally a damping effect, can actually cause the instability; no student of fluid mechanics should regard his education as complete until he understands how this happens. The analysis of other related flows, including the important practical case of a boundary layer on a flat plate, follows naturally from the discussion of the second of these two prototypes.

There is much to be learnt from this little book about the progress of research in recent years. It is clearly and carefully written, and seems to provide a suitable reference book to accompany a lecture course on advanced fluid mechanics.

G. K. BATCHELOR.

CONTENTS OF SECTION A

PAGE

Dr. D. T. Stewart. Electron Excitation Functions of the First Negative Bands of N_2^+ 437

Dr. S. Nakajima. A Note on Momentum Correlation and Diamagnetism . . 441

Dr. D. M. Thomson. The Yield and Angular Distribution of Neutrons from the Reaction ^{63}Cu(p, n)^{63}Zn at 18 mev 447

Dr. E. Breitenberger. Angular Correlations between Successive Conversion Electrons in ^{80}Br 453

Mr. C. H. Collie, Dr. R. E. Meads and Mr. E. E. Lockett. The Capture Cross Section of Neutrons by Boron 464

Mr. R. E. Meads, Mr. C. J. England, Mr. C. H. Collie and Mr. G. C. Weeks. The Capture Cross Section of Thermal Neutrons in Water 469

Dr. B. L. Moiseiwitsch and Dr. A. L. Stewart. Approximate Molecular Orbitals: V—The $1s\sigma$ and $2p\pi$ states of HeH^{2+} 480

Research Notes:

- Dr. E. Breitenberger. Remarks on the Least-Squares Reduction of Angular Correlation Data 489
- Dr. C. G. Kuper. On the Bohm–Pines Theory of a Quantum-Mechanical Electron Plasma 492
- Dr. J. M. Cassels. Proton–Proton Scattering at 147 mev 495
- Mr. B. Brown, Dr. E. Matsukawa and Prof. E. A. Stewardson. The Nuclear Scattering of 0·4 and 0·5 mev Electrons in Composite Aluminium–Uranium Foils 496

Reviews of Books 500

Contents of Section B 504

Figure 3. Bi_2Te_3 prepared by directional freezing in a horizontal boat. The ripple marks indicate the shape of the solid–liquid interface during freezing. Stray crystals may be seen growing in from the side of the boat towards the last end of the ingot to freeze.

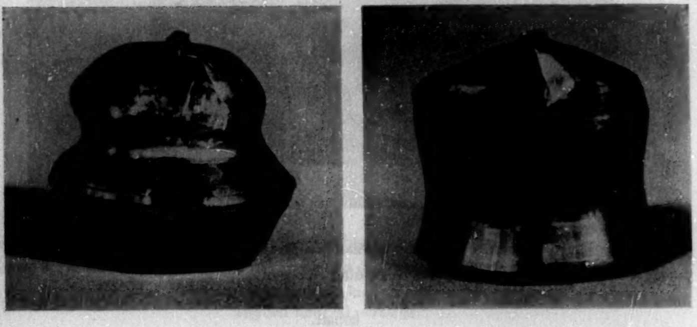
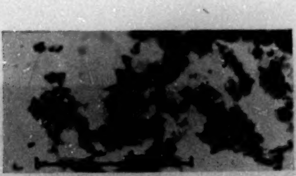
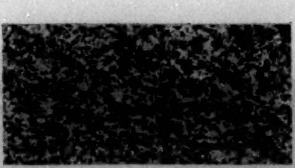


Figure 6. Bi_2Te_3 ingots grown by pulling from the melt. The stumps of the seeds are 3 mm in width.

(1)

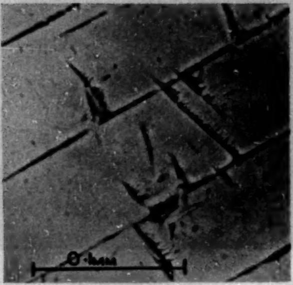


(2)



(3)

(4)



(5)

Figure 1. Dried magnetite colloid of normal composition.

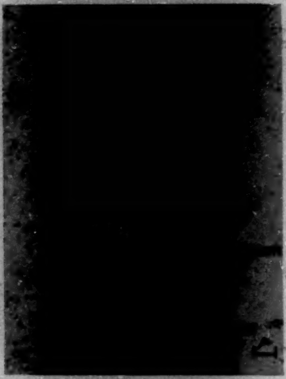
Figure 2. Dried magnetite colloid containing ‘ Celacol ’.

Figure 3. Phase-contrast micrograph of ‘ Celacol ’-treated colloid on hexagonal face of cobalt.

Figure 4. Phase-contrast micrograph of ‘ Celacol ’-treated colloid on (1000) face of cobalt showing 180° walls.

Figure 5. Electron micrograph of treated colloid on hexagonal face of cobalt.

(6)



(7)

(8)

(9)

Figure 6. Portion of the pattern of figure 5 at higher magnification.

Figure 7. Portion of an electron micrograph of a dagger closure domain on a cobalt crystal.

Figure 8. Portion of an electron micrograph of a lace pattern on (110) surface of silicon–iron; applied field 40 oersteds.

Figure 9. Electron micrograph of 180° wall with striation deposits on (100) face of silicon–iron.

(10)

(11)

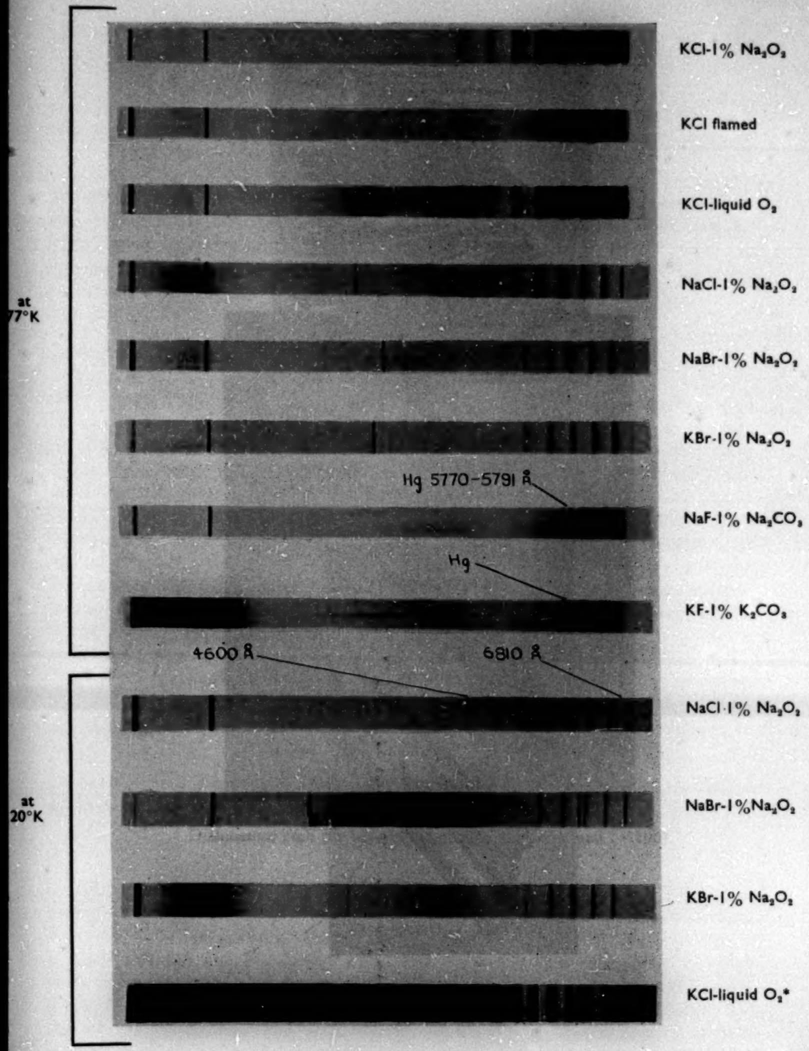
(12)

(13)

Figure 10. Electron micrograph of 180° wall on (100) face of silicon–iron.

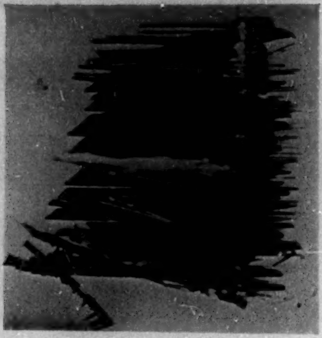
Figures 11, 12, 13. Electron micrographs of 180° boundaries obtained with successively more dilute colloids, the relative dilution for neighbouring photographs being approximately 1 : 4.

PROC. PHYS. SOC. VOL. 69, PT. 6—B (J. EWLES AND D. S. BARMBY).

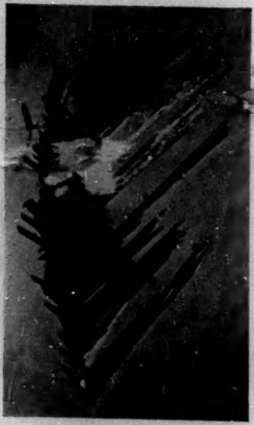

KCl-1% Na_2O_2
KCl flamed
KCl-liquid O_2
NaCl-1% Na_2O_2
NaBr-1% Na_2O_2
KBr-1% Na_2O_2
Hg 5770–5791 Å
NaF-1% Na_2CO_3
Hg
KF-1% K_2CO_3
4600 Å
6810 Å
NaCl-1% Na_2O_2
NaBr-1% Na_2O_2
KBr-1% Na_2O_2
KCl-liquid O_2*
at
77°K
at
20°K

*Enlarged 1½ times others

Luminescent emission spectra of some oxygen activated alkali halide phosphors. Spectrograph—Hilger E2, Slit width 0·05 mm, Plates—Kodak spectrographic 1F, Exposures 10 to 60 minutes.

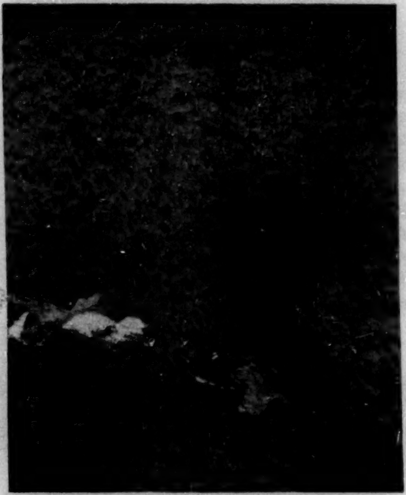


(a)



(b)

Figure 1. Crystals of hexagonal selenium. Scale 3 cm = 1 mm.



Dislocation etch pits on (111) face of a silicon crystal (×100).

Carrier Accumulation in Germanium

By J. B. ARTHUR, A. F. GIBSON and J. B. GUNN
Radar Research Establishment, Malvern, Worcs.

MS. received 11*th October* 1955, *and in revised form* 18*th February* 1956

Abstract. Contacts can be made to germanium crystals which are largely impermeable to minority carriers but permeable to majority carriers. In the presence of a suitable electric field high concentrations of minority carriers can accumulate at such contacts. It is the purpose of this paper to describe some experiments which demonstrate the existence of this effect.

---

§ 1. Introduction

It is well known that the entry of minority carriers into a normally doped germanium crystal can be markedly reduced by the interposition of a heavily doped section of crystal of the same conductivity type between the main body of the crystal and the external contact. Thus, for example, Arthur *et al.* (1955) used a heavily doped n-type region (designated n+) in conjunction with lightly doped n material to exclude the entry of holes from the positively biased contact. Similarly it might be expected that such an n–n+ junction would restrict the exit at the negative terminal of holes from normally doped n-type material. Hence the conditions required by Low (1955) for the observation of the minority carrier accumulation effect would be achieved and a high concentration of holes at the junction expected.

The purpose of this paper is:

(1) To discuss briefly and qualitatively the theory of n–n+ and similar junctions and to indicate the conditions necessary to observe minority carrier accumulation.

(2) To describe some direct experimental observations of carrier accumulation.

(3) To introduce the current gain effect which is a consequence of accumulation and the subject of the succeeding paper.

§ 2. The L–H Junction

For greater generality we shall designate junctions of the n–n+ type L–H junctions where L and H refer to light and heavy doping concentrations respectively. The arrangement to be discussed consists of a filament containing an L–H junction at right angles to its length. The H and L sections are assumed to be of infinite length (i.e. long compared with the downstream diffusion length) a condition which can be achieved to a good approximation experimentally by the use of 'dumbbell' end connections. A sweeping field is applied with a polarity such as to cause minority carriers in the L region to flow towards the L–H junction. If, as indicated above, the L–H junction is largely impermeable to minority carriers but not to majority carriers, a high concentration of minority carriers will accumulate at the L side of the L–H junction.

The low permeability of the L–H junction to minority carriers arises essentially from the different modes of carrier transport on the two sides of the junction. Because of the low resistivity of the H region, the electric field there is negligible and the carriers flow only by diffusion. The passage of a minority carrier current across the junction will therefore cause a concentration gradient to be set up which will, in turn, give rise to a large concentration of carriers at the junction if the lifetime in the H region is not too small.

To put the above argument on a more quantitative basis and to indicate the experimental conditions required to observe accumulation we shall show that the distribution of minority carriers is that shown in figure 1 which also serves to define the symbols used. For greater generality we may suppose that there is an additional source of carriers (e.g. an emitter contact) on the L side.

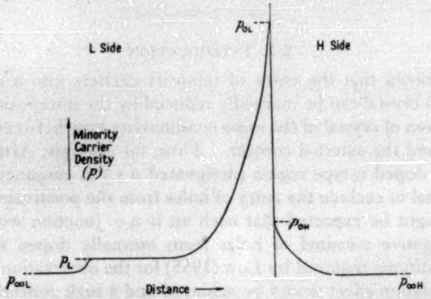


Figure 1. Expected minority carrier distribution in a biased L–H junction.

The hole current density in the H region of an n-type L–H junction may be calculated on the same basis as the hole saturation current of a p–n junction and is

$$i_p = q(p_{0H} - p_{\infty H})\sqrt{\frac{D_H}{\tau_H}} \qquad \ldots\ldots(1)$$

where D_H and τ_H are the diffusion constant and lifetime of the minority carriers on the H side. If the junction is abrupt and the voltage drop across it small compared with $\boldsymbol{k}T/q$ the electron and hole quasi-Fermi levels will be unchanged through the junction so that

$$\frac{p_{0H}}{p_{0L}} = \frac{n_{0L}}{n_{0H}} = \exp(-\phi/\boldsymbol{k}T) \simeq \frac{\sigma_L}{\sigma_H} = \frac{p_{\infty H}}{p_{\infty L}} \qquad \ldots\ldots(2)$$

where ϕ is the difference in hole potential on each side of the junction and the other symbols have their usual meaning. Equation (2) can be modified very simply to allow for a finite voltage drop across the junction by reducing the value of ϕ.

In the absence of recombination in the very narrow junction region the hole current must be the same on both sides and hence

$$i_p = q(p_{0L} - p_{\infty L})\frac{n_{0L}}{n_{0H}}\sqrt{\frac{D_H}{\tau_H}} = q(p_{0L} - p_{\infty L})V_{0L} \qquad \ldots\ldots(3)$$

where V_{0L} may be called the leakage velocity of the junction. As the hole current i_p is determined by the field conditions and the hole density p_L in the L region far from the junction, the concentration of accumulated holes p_{0L} can become very large if V_{0L} is made small.

The important properties of an L–H junction may therefore be tabulated as follows:

(i) The leakage velocity of minority carriers across the junction may be made very low if the resistivity ratio is large and an adequate lifetime is maintained in the heavily doped section.

(ii) The leakage velocity is independent of the sweep current provided no significant electric field exists in the H region. This condition is satisfied in all the experiments described. It may be noticed that this result makes the current ratio γ, as defined by Low (1955) for example, an unsuitable parameter for describing the properties of L–H junctions as γ will be a function of sweep current at small currents.

(iii) Equation (3) indicates that the leakage velocity will increase when the minority carrier concentration is comparable with, and causes a significant increase in, the majority carrier concentration n_{0L}. In addition enhanced recombination in the accumulation region may be expected under such conditions. Hence the density of accumulated holes may be expected to saturate with time at some given value but the experimental results cannot indicate whether leakage or recombination is the dominant process.

The above theoretical discussion is a much simplified version of a comprehensive quantitative theory of L–H junctions developed by Gunn which, it is hoped, will be submitted for publication shortly. As all the experiments to be described are, however, of a qualitative nature the above theory is sufficient for present purposes and may serve as an introduction to Gunn's paper.

§ 3. EXPERIMENTAL DETAILS

Minority carrier accumulation can and does occur in systems other than L–H junctions, for example on the low recombination velocity surface of a magnetic rectifier of the type described by Welker (1953). The present discussion will, however, be limited to L–H junctions. The simplest method, from the point of view of experimentation and interpretation, of demonstrating carrier accumulation directly is illustrated in figure 2. The measurement of current multiplication, to be described in the succeeding paper, represents an indirect observation of accumulation.

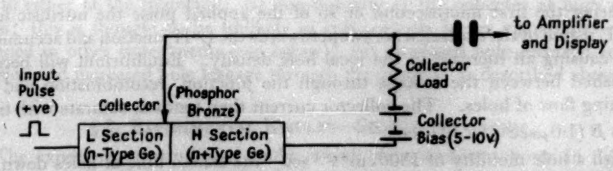


Figure 2. Experimental arrangement for the observation of carrier accumulation.

The experimental arrangement (shown in figure 2) consists of an L–H filament with a suitable point contact collector contact (formed or unformed) placed at various points on the L region.

A small bias is applied to the collector contact. The collector current signal is amplified in a wide bandwidth amplifier (7 Mc/s bandwidth) and displayed on an oscilloscope. It is assumed that the collector current is a monotonically increasing function of local hole density and hence the latter can be observed as a function of time. A voltage pulse, typically of 10 microseconds duration and variable amplitude, is applied to the rod containing the L–H junction so that minority carriers accumulate at the junction. A pre-pulse, about 0·75 microseconds before the main pulse, initiates the start of the oscilloscope time base.

In all the experiments described in the present section n-type L–H junctions were used and all were grown in a conventional crystal pulling equipment by Dr. W. Bardsley. The resistivities of the n sections ranged from 8 to 40 ohm cm and the heavily doped n sections had resistivities of the order of 0·002 ohm cm.

A great variety of collector current waveforms, some of considerable complexity, can be obtained under the above experimental conditions depending, for example, on the nature of the end contact to the L section and the intensity of any ambient illumination. Essentially similar results were obtained with all the junctions examined and hence to facilitate description the waveforms obtained on only one specimen will be described. The resistivity and lifetime of the L section of this specimen were 8 ohm cm and 50 microseconds respectively, the rod having a cross-sectional area of about 25×10^{-4} cm^2 and the L section being 0·32 cm long. The H section was about 0·7 cm long. Both 'dumbbell' type and soldered connections were used on the L section, the marked amount of injection occurring in the latter representing a useful source of additional minority carriers to accumulate at the junction.

Some typical collector signal waveforms, obtained under various conditions, are shown in figures 3, 4 and 5. The experimental conditions and suggested interpretation of each waveform are given in the following sections.

§ 4. Experimental Results—First Waveform

The experimental conditions under which the waveform shown in figure 3 was obtained were as follows: (*a*) collector contact about 0·1 mm or less from the L–H junction, (*b*) no external illumination, (*c*) applied pulse amplitude about 50 volts, duration 10 microseconds. The waveform can be conveniently split into three regions, designated A, B and C on the diagram. The suggested interpretation of this waveform is as follows:

Region A ($0 < t < 1{\cdot}0\,\mu$sec).

During the first microsecond or so of the applied pulse the intrinsic holes in the L region of the specimen are swept towards the L–H junction and accumulate there, causing an increase in the local hole density. Equilibrium will become established between the leakage through the junction, recombination and the oncoming flow of holes. The collector current then tends to saturate with time.

Region B ($1{\cdot}0\,\mu\text{sec} < t < 10\,\mu$sec).

With a hole mobility of 1800 cm^2 v^{-1} sec^{-1} the transit time of holes down the L section with 50 volts applied should be 1·1 microseconds. At or about this time after the pulse is switched on, holes at the positive-going soldered contact begin to arrive and accumulate at the L–H junction. Again an equilibrium situation is reached after a few microseconds. Under the conditions of the experiment, however, the density of accumulated holes may now be comparable with the original electron density. If so, the leakage velocity through the junction and the recombination rate of holes will be increased. As the equilibrium hole density in 8 ohm cm material is about a tenth of the equilibrium electron density, and as the signal amplitude in region B is about five times that in region A, which in turn probably corresponds to a hole density markedly in excess of the equilibrium value, it is likely that the above condition is achieved.

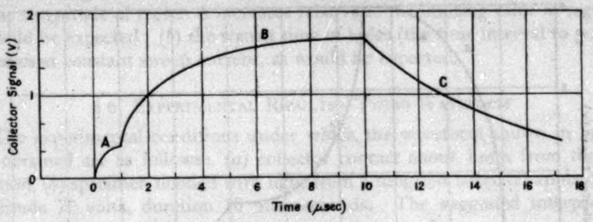


Figure 3. Accumulation waveform with collector very near junction.

Region C ($t > 10\,\mu$sec).

When the applied field is switched off, the density of accumulated holes decreases rapidly due to recombination with electrons and diffusion back into the body of the specimen. The form of the decay curve observed in this region is critically dependent on the position of the collector contact relative to the L–H junction and the simple decay indicated in figure 3 is obtained only when the collector is very close to the junction. This result will be discussed more fully in § 6.

4.1. *The Effect of Electric Field Variation*

As the magnitude of the applied sweep current pulse is increased the following effects are observed: (*a*) The time interval to the cusp AB, which represents the transit time of injected holes down the length of the L region, decreases monotonically as would be expected. (*b*) The signal amplitude in region B (and, therefore, C) increases as more holes are injected due to the increasing sweep current. The density of accumulated holes does not increase indefinitely, however, but tends to saturate at high sweep currents. This saturation may be due to either of the following: (i) non-linearity of the collector at high signal levels: this is thought to be unlikely as substantially larger signals have been obtained in other experiments (see below); (ii) increased hole–electron recombination and increased hole leakage through the junction as described above.

§ 5. Experimental Results—Second Waveform

The experimental conditions under which the waveform shown in figure 4 was obtained were as follows: (*a*) collector contact about 0·1 mm or less from the L–H junction, (*b*) specimen flooded with light from a tungsten lamp, (*c*) applied pulse amplitude about 20 volts, duration 10 microseconds.

Region A ($0 < t < 5\,\mu$sec).

During the relatively long period between the applied pulses (typically 2–10 milliseconds), the background light generates a large density of holes in the L region. On the application of the field these holes are swept along the rod to accumulate at the L–H junction. The light is however unable to maintain this high density of holes in the L region during the application of the pulse (cf. Lawrance and Gibson 1952). This source of holes is therefore exhausted after a time equal to the transit time down the filament. This occurs at point B.

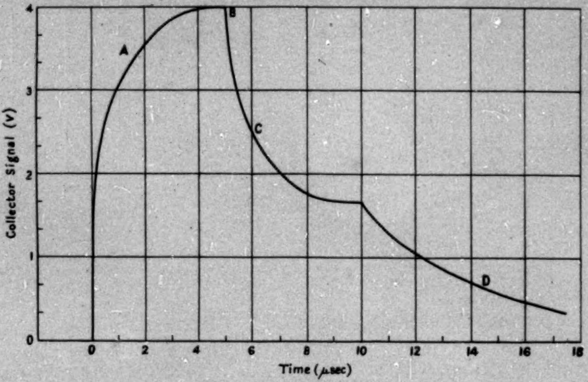


Figure 4. Waveform in presence of bright background light.

Region C ($5\,\mu$sec $< t < 10\,\mu$sec).

In spite of the continued application of the field there is now only an insignificant number of optically generated holes to accumulate at the junction. Therefore the quasi-equilibrium condition reached at point B cannot be maintained and the density of accumulated holes falls rapidly, presumably due to recombination and leakage through the junction. The density of accumulated holes represented by point B almost certainly exceeds the initial majority carrier density, as indicated by the magnitude of the signals obtained (cf. figure 3, region B).

Ultimately a new equilibrium state is reached having a magnitude comparable with that of region B, figure 3, and invariant with background light intensity. The dominant source of holes is now end injection, as before.

Region D ($t > 10\,\mu$sec).

Finally the applied field is switched off and the density of accumulated holes decays away.

5.1. *The Effect of Electric Field and Light Intensity Variation*

As the magnitude of the applied pulse is increased, the following effects are observed: (*a*) The time interval to the point B, which represents the transit time of holes along the filament, decreases monotonically. The apparent mobility is, however, less than the normal value as the large density of holes created by the light is comparable with the electron density (cf. Arthur *et al.* 1955). (*b*) The magnitude of the whole waveform increases and very large signals (10–12 volts in amplitude) can be obtained. The signal amplitude tends to saturate at high field due to increased leakage through the junction and enhanced recombination. (*c*) The rate of decay in region C increases, indicative of enhanced leakage and recombination. The initial decay, covering half the total drop in amplitude, becomes very rapid indeed and is limited by the bandwidth of the display equipment.

As the background light intensity is increased the following effects are observed: (*a*) the magnitude of region A increases relative to the limiting value of region C, as would be expected; (*b*) the transit time of holes (the time interval to point B) increases at constant sweep current, as would be expected.

§ 6. EXPERIMENTAL RESULTS—THIRD WAVEFORM

The experimental conditions under which the waveform shown in figure 5 was obtained are as follows: (*a*) collector contact about 1 mm from the L–H junction, (*b*) specimen flooded with light from a tungsten lamp, (*c*) applied pulse amplitude 70 volts, duration 10 microseconds. The suggested interpretation of this waveform is as follows:

Regions A and B ($0<t<10\,\mu$sec).

The important aspect of this arrangement is that the collector is placed a considerable distance from the L–H junction. One result of this is an appreciable 'ohmic' component of the collector signal due to the voltage drop along the filament during the application of the pulse.

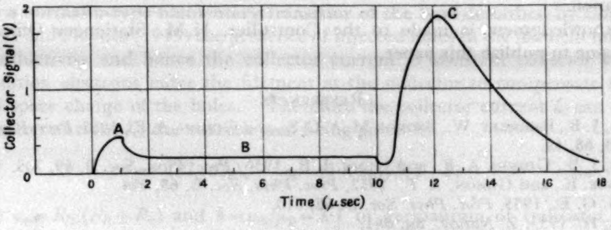


Figure 5. Waveform with collector far from junction.

A noticeable increase in hole density at the collector does occur during the first microsecond of the applied pulse (region A). As the accumulation of holes at the junction proceeds the hole density at the collector decreases to a steady, low, value. Thus this portion of the waveform is indicative of the rate of accumulation of holes at the L–H junction. Initially, at zero time the hole density is everywhere uniform and as accumulation proceeds the spatial distribution of holes on the L side of the junction becomes progressively higher and steeper. The local hole density at a point far from the junction can therefore rise and fall as observed. This type of behaviour is observed to a greater or lesser extent at all points in the L region except very close to the L–H junction.

Region C ($t > 10\,\mu$sec).

When the applied pulse ceases a high concentration of holes has accumulated at the L–H junction. In the absence of the applied field this concentration gradient causes diffusion of carriers back towards the collector contact. Initially the collector current falls, due to the drop in ohmic signal, but subsequently the holes diffuse past the collector causing a marked rise and then fall in collector current.

The above phenomenon, which will be referred to as the 'rebound' effect, can be observed at almost any position of the collector except very close to the L–H junction. Indeed a simple decay of the type indicated by figures 3 and 4 requires very careful placing of the collector point. In general the collector waveform is a complex function of a simple decay and a rebound pulse.

§ 7. CONCLUSIONS

The leakage velocity of minority carriers through an L–H junction can have a very small value in suitably designed junctions. If the leakage velocity is less than the drift velocity in the L region far from the junction, minority carriers must necessarily accumulate at such junctions. It is believed that the experiments described, and particularly the occurrence of the rebound effect, provide convincing and direct evidence of the existence of this effect. In the following paper (Arthur, Gibson and Gunn 1956) current gain which is a necessary consequence of accumulation will be described.

ACKNOWLEDGMENTS

We are indebted to our colleagues at this Establishment for their advice and help. We also wish to thank Mr. Low for allowing us to see his paper prior to publication.

Acknowledgment is made to the Controller, H.M. Stationery Office, for permission to publish this paper.

REFERENCES

ARTHUR, J. B., BARDSLEY, W., BROWN, M. A. C. S., and GIBSON, A. F., 1955, *Proc. Phys. Soc.* B, **68,** 43.

ARTHUR, J. B., GIBSON, A. F., and GUNN, J. B., 1956, *Proc. Phys. Soc.* B, **69,** 705.

LAWRANCE, R., and GIBSON, A. F., 1952, *Proc. Phys. Soc.* B, **65,** 994.

LOW, G. G. E., 1955, *Proc. Phys. Soc.* B, **68,** 310.

WELKER, H., 1953, Z. *Naturf.*, **8a,** 681.

Current Gain at L–H Junctions in Germanium

By J. B. ARTHUR, A. F. GIBSON and J. B. GUNN
Radar Research Establishment, Malvern, Worcs.

MS. received 11*th October* 1955, *and in revised form* 18*th February* 1956

Abstract. Junctions between lightly and heavily doped germanium of the same conductivity type (L–H junctions) have a low leakage velocity for minority carriers. This result, which is discussed in the preceding paper, can be considered as a reduction of the minority carrier mobility and gives rise to enhanced current gain in certain structures, for example filamentary transistors. The paper describes experiments designed to demonstrate the existence of this effect and to indicate its magnitude.

---

§ 1. Introduction

The accumulation of minority carriers which occurs at biased L–H junctions has been described in the preceding paper (Arthur, Gibson and Gunn 1956) and familiarity with this paper will be assumed. It is possible to show, by simple physical arguments, that current gain must also occur. The detailed theory of this effect has been calculated by Gunn (to be published) but, as before, we shall use only qualitative arguments to predict the experimental results.

In a normal n-type filamentary transistor of the type described by Shockley, Pearson and Haynes (1949) the presence of injected holes in the filament increases its conductivity and hence the collector current at constant collector voltage. In addition electrons enter the filament at the collector to compensate for the added space charge of the holes. Therefore the collector current I_c can exceed the emitter current I_e, the current gain being given by

$$\alpha=(dI_c/dI_e)_{V_c}=\alpha_0+(1-\alpha_0)\gamma\beta(1+b) \qquad \ldots\ldots(1)$$

where $\alpha_0=R_b/(R_b+R_c)$ and $b=\mu_n/\mu_p=2{\cdot}1$ in germanium of transistor purity at room temperature. μ_n and μ_p are the mobilities of electrons and holes respectively, γ and β constants of the order of but less than unity and R_c, R_b the collector and base resistance respectively. The maximum value of α is therefore $1+b$ and about $3{\cdot}1$ under normal conditions.

Consider now a filamentary transistor containing an L–H junction, as illustrated in figure 1. We shall show that, if appreciable accumulation of minority carriers occurs at the L–H junction, the current gain of the transistor can be greater than that indicated by equation (1) and often greater than $1+b$. The large current gain arises because the accumulation of a high density of holes increases the local conductivity by a very large amount and allows a large electron current to flow. The L–H junction may be considered as a device for reducing the hole mobility, at least locally, with a consequent increase in the effective mobility ratio b.

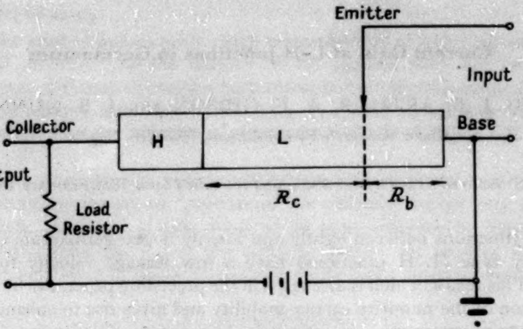


Figure 1.—Schematic diagram of an n-type filamentary transistor containing an L–H junction. Resistance of H region assumed negligible.

§ 2. THE OBSERVATION OF CURRENT GAIN IN FILAMENTARY TRANSISTORS CONTAINING L–H JUNCTIONS

Clearly the most important experiment that can be carried out is a direct observation of current gain in an arrangement of the type shown in figure 1. In order to demonstrate convincingly that the junction enhances the current gain, however, certain conditions are at least desirable if not essential, as follows.

If at all possible a current gain greater than 3·1 ($=b+1$) should be obtained at room temperature as this represents the maximum achievable value in a normal filamentary structure. The important factors are:

(*a*) The value of α_0. Clearly from equation (1) the current gain will be a maximum where R_b is small compared with R_c.

(*b*) The distance of the emitter from the L–H junction. Even if the minority carrier lifetime in the filament is infinite so that no carriers are lost in transit through the L region and the condition $R_b \ll R_c$ is achieved the current gain will increase as the emitter is brought nearer to the L–H junction. This arises because the portion of the L region in which the holes accumulate is very short (~0·1 mm or less) and only in this portion is the modulation of filament conductivity large. If the emitter is placed a considerable distance from the L–H junction most of the filament (and hence R_c) is only modulated to a depth indicated by equation (1).

(*c*) The nature of the base connection. Equation (1) assumes that the base connection is non-injecting. If however minority carriers can be injected by the base connection an enhanced value of α can be obtained due to this effect alone. This arises because injection of holes from the emitter gives rise to an increase in base current which injects further holes, which increases the base current still further, and so on. A suitable base contact therefore is the 'dumbbell' type which is equivalent to making the filament infinitely long (Shockley, Pearson and Haynes 1949).

(*d*) As the leakage velocity of holes through the junction and the recombination rate near the junction will increase when the density of accumulated holes becomes large it is to be expected that the current gain will decrease at high emitter currents. Hence the experiment is likely to be most successful at low emitter currents.

With these aspects in mind the arrangement adopted experimentally was that shown in figure 2. The resistivity of the L region of the specimen was 8 ohm cm and its cross-sectional dimensions were about $0{\cdot}4 \times 0{\cdot}4$ mm^2. It was found necessary to place the emitter less than 0·4 mm from the collector contact in order to obtain a large current gain. This destroys the one-dimensional aspect of the experiment but fortunately does not invalidate the main purpose of the experiment as the current gain still cannot exceed (and indeed would be less than) that indicated by equation (1) unless increased by the L–H junction.

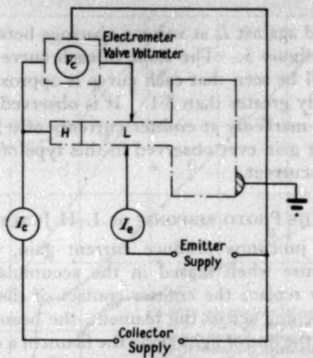


Figure 2.—Schematic diagram of circuit used to measure current gain in a filamentary transistor.

In order to reduce the effective value of R_b to a minimum a voltage probe was put close behind the emitter contact and the collector voltage measured

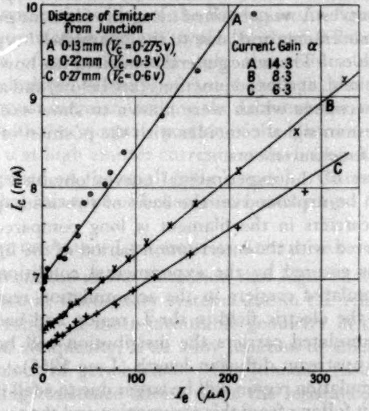


Figure 3.—Collector current (I_c) vs. emitter current (I_e) for various emitter positions.

relative to this probe. The collector voltage was measured with the aid of an electrometer valve voltmeter of the type described by Scroggie (1952) which has an input impedance of the order of 10^{13} ohms. The experiment then consisted of varying I_e, adjusting the collector supply to give a constant predetermined value of V_c and observing I_c.

Curves of I_c plotted against I_e at various spacings between emitter and L–H junction are shown in figure 3. The slope of each curve gives the value of the current gain α. It will be seen that each curve is approximately linear and the current gain appreciably greater than 3·1. It is observed, as expected, that the current gain decreases markedly at emitter currents of 0·5 mA or greater. The largest value of current gain ever observed in this type of experiment was 30 at small values of emitter current.

§ 3. THE PHOTO-RESPONSE OF L–H JUNCTIONS

Inasmuch as L–H junctions produce current gain, they show a marked photoconductive response when biased in the accumulation direction. It is therefore of interest to replace the emitter contact of the previous experiment by a line of light, stretching across the filament, the beam being interrupted at 800 c/s. By traversing the line of light along the filament a measure of the current gain can be obtained as a function of distance without resort to the somewhat artificial use of a voltage probe.

To be strictly analogous to the previous experiment the photo signal should be measured with a constant voltage across the specimen. As is common in this type of work, however, the signal under constant current conditions is simpler to interpret. In the absence of any enhanced gain at the L–H junction the signal will be proportional to the distance between the light spot and the junction (the H region being assumed to be of negligible resistance). Any multiplication effects at the junction will be indicated by a signal near the junction greater than that expected from a linear increase of signal with distance.

Multiplication effects are clearly indicated by the set of typical curves given in figure 4. Of these curves A was obtained using the filament previously described for current gain measurements and, due to the 'dumbbell' type end connection, may be the most reliable. The same general behaviour has however been obtained for all junctions examined, at alloyed junctions (see below) and at soldered contacts to high resistivity specimens which were known to show exclusion effects. In each case the rapid rise in signal coincides with the position of the L–H junction to within the limits of measurements.

The main features of the experimental curves obtained in the absence of background light can be explained on the basis of two assumptions, as follows: (*a*) The lifetime of carriers in the filament is long compared with the transit time but short compared with the interruption period of the light. The validity of this assumption is assured by the experimental conditions used. (*b*) The distribution of accumulated carriers in the accumulation region is determined by the magnitude of the electric field in the L region and back diffusion. For low densities of accumulated carriers the distribution will be exponential and characterized by the upstream diffusion length (Low 1955); for high densities the width of the accumulation region will be larger due to a fall in the local electric field. In either case it follows from this assumption and the previous one that the distribution of carriers in the accumulation region is independent of the position of the light spot in the L region.

The observed variation in signal with light position then arises as follows. Illumination in the H region produces virtually no signal due to the already low resistivity of this region. Illumination in the accumulation region leads to a large signal which remains substantially constant throughout the accumulation region. Illumination in the bulk of the L region leaves the signal from the accumulation region unchanged and in addition modulates the resistivity of the rod between the light spot and the edge of the accumulation region. The latter contribution increases linearly with distance, as observed.

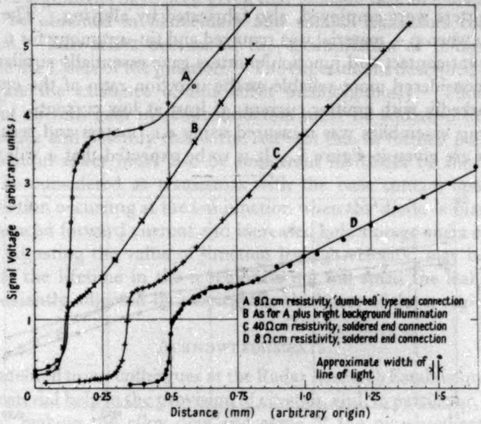


Figure 4.

On this interpretation the width of the flat regions in figure 4 is a measure of the width of the accumulation regions, typically about 0·2 mm.

In the presence of a background light the results are inevitably more complicated. The leakage velocity of holes through the junction and their recombination rate must increase, making the accumulation region relatively less effective compared with the 'normal' portion of the filament. This is essentially similar to the fall in α at high emitter currents as the background light represents a d.c. bias superimposed on the small 800 c/s signal. Curve B of figure 4 does appear to indicate that this type of behaviour is occurring.

§ 4. THE USE OF ALLOY JUNCTIONS AS ACCUMULATING CONTACTS

It will be clear from the above discussion that high values of current gain α are difficult to achieve in filamentary systems due to the relatively high base to collector resistance ratio. A more favourable geometry in this respect, though of greater theoretical complexity, is an alloy junction of small area made on the surface of a large block of germanium. The junction (collector) resistance is large due to spreading resistance effects while the resistance of the bulk of the material (base resistance) remains low.

With the above considerations in mind the current gain occurring at L–H junctions formed by alloying on both n-type and p-type material was measured. The base contact was also an alloyed L–H junction of large area to exclude the possibility of spuriously large values of α being obtained due to base injection (see above). Initially point contact emitters were used but subsequently p–n junction emitters were employed, also fabricated by alloying. The alloys used were indium when p+ material was required and tin–antimony for n+ material.

While point contact and junction emitters gave essentially similar results the latter were considered more reliable as the injection ratio of the emitter would not vary markedly with emitter current, at least at low currents. The current gain of various assemblies was measured using d.c. meters and results on three typical units are given in figure 5. It is to be expected that α will fall off with increasing emitter current due to increased recombination at the junction and leakage through the junction. In spite of this the current gain remains well above $b+1$ over a substantial range.

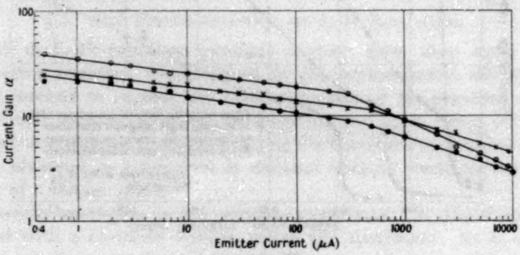


Figure 5.

§ 5. CURRENT MULTIPLICATION AT FORMED POINT CONTACT COLLECTORS

It has been known for many years that electro-formed collector contacts on n-type germanium can give a current gain in excess of unity if the contact wire contains a significant n-type impurity—for example phosphor-bronze wire. To explain this result Valdes (1952) has assumed the formation of a p–n hook at the collector, the p material being produced by thermal conversion and the n region by injection of impurities from the contact wire. An alternative model due to Sittner (1952) invokes traps at the collector and does not indicate why the composition of the wire should be important.

It has recently been shown that similar forming rules apply to silicon point contact collectors (Granville *et al.* 1955). Wires containing a group V element must be used for n-type silicon and a group III element for p material. While thermal conversion effects of a complex nature do occur in some samples of silicon (Fuller *et al.* 1955) it is not possible to invoke the p–n hook mechanism in both n and p material as thermal conversion at a given forming temperature is a one way process.

To explain the above results, therefore, it is now suggested that electro-forming results in all cases in the formation of an L–H junction. It is known that rectification either deteriorates or reverses in direction on forming (Granville *et al.*

1955), a result consistent with this view. A high contact resistance is maintained due to effects of the type described by Gunn (1952).

§ 6. CONCLUSIONS

Qualitative arguments have been given that indicate that an enhanced current gain should be obtained in filamentary and similar transistors containing an L–H junction. This arises because of the accumulation of minority carriers that occurs on the L side of the junction. The experiments described are believed to demonstrate the existence of this effect unambiguously. Thus one more mechanism of current gain in semiconductors must be added to the p–n hook and trap theories and possibly this is the relevant one in formed point contacts. The accumulation effect must also be important in diodes of the p–i–n type which may be considered as transistors with the base contact open-circuited. The accumulation occurring at the i–n junction when the diode is biased forward will give enhanced forward current and increased hole storage and a compromise, achieved by adjusting the value of junction leakage velocity, may be necessary. Provided that the lifetime in the n region is not too small the leakage velocity can be conveniently adjusted by varying the thickness of the n region.

ACKNOWLEDGMENTS

We are indebted to our colleagues at the Radar Research Establishment for their advice and material help in the provision of crystals, and, in particular, to Dr. J. W. Granville for making the alloy dots and some of the measurements discussed in §4.

Acknowledgment is made to the Controller, H.M. Stationery Office, for permission to publish this paper.

REFERENCES

ARTHUR, J. B., GIBSON, A. F., and GUNN, J. B., 1956, *Proc. Phys. Soc.* B, **69**, 697.

FULLER, C. S., DITZENBERGER, J. A., HANNAY, N. B., and BUEHLER, E., 1955, *Acta. Met.*, **3**, 97.

GRANVILLE, J. W., BARDSLEY, W., and GIBSON, A. F., 1955, *Brit. J. Appl. Phys.*, **6**, 206.

GUNN, J. B., 1952, *Proc. Phys. Soc.* B, **65**, 908.

LOW, G. G. E., 1955, *Proc. Phys. Soc.* B, **68**, 310.

SCROGGIE, M. G., 1952, *Wireless World*, January 1952, p. 14.

SHOCKLEY, W., PEARSON, G. L., and HAYNES, J. R., 1949, *Bell Syst. Tech. J.*, **28**, 364.

SITTNER, W. R., 1952, *Proc. Instn Radio Engrs, N.Y.*, **40**, 448.

VALDES, L. B., 1952, *Proc. Instn Radio Engrs, N.Y.*, **40**, 445.

Current Gain in Formed Point Contact n-type Germanium Transistors

By D. HANEMAN†

Division of Radiophysics, Commonwealth Scientific and Industrial Research Organization, Sydney

Communicated by E. G. Bowen ; MS. received 27th September 1955, *and in revised form 19th December* 1955

Abstract. Experiments on germanium point contact transistors in which collector points of pure donor and acceptor elements are formed show that in general donors appear to be essential to obtain current gain improvements. Successful forming was also achieved with pure gold points. It is assumed that the atoms of the material of the collector diffuse into the germanium during forming. This leads to a change in the amount of lowering of the surface barrier due to the holes injected at the emitter, and accounts for the increased current gain.

§ 1. Introduction

Two main lines of approach have been used in attempts to explain the phenomenon of current multiplication at the formed collector of an n-type germanium point contact transistor. Both theories are based on the concept of injected holes from the emitter being delayed in front of the collector, but neither discussion is free from serious objection. Direct experimental evidence of the physical changes occurring at the collector point during forming as used for transistors has been difficult to obtain.

In the present work the effect of forming on the gain has been studied for collectors of several materials, including pure donors. It appears that infused donor atoms enhance the barrier lowering, which arises from the injected holes from the emitter, as discussed by Hogarth (1954).

§ 2. Theories of Current Multiplication

It was suggested by Shockley (1950) that the forming process may introduce hole traps in the germanium immediately in front of the collector. A theoretical discussion of this concept has been given by Sittner (1952) who used the theory of traps as developed for filamentary transistors. However Hogarth (1954) has shown experimentally that the hole trap theory of Sittner is not applicable to point contact transistors. Further, observed values of α considerably greater than 3·1 predicted by Sittner at emitter current I_e of order milliamps are not readily explicable on this approach.

Shockley (1950) has also suggested the possibility of a p–n hook appearing in front of a formed point (figure 1). According to the theory of Shockley, Sparks and Teal (1951), the intrinsic current gain α^* of a pnpn structure is $(1+\sigma_n L_n/\sigma_p W)$ where σ_n/σ_p is the conductivity ratio of the n and p regions in front of the collector, L_n the diffusion length for holes in the collector n region and W the thickness of the collector p region. For small values of W high values of α^* may be obtained.

---

†Now at the Department of Physics, University of Reading.

The serious objection that the high forward resistance expected from this structure is not observed with a formed collector can be met by postulating high values of σ_n and σ_p. Then the junction nearest the positive biased collector will have a very high potential gradient, being in reverse bias, and will break down at low values of applied voltage. The other junction will have a high breakdown voltage since most of the potential drop will be in the comparatively high resistivity n-type germanium. A breakdown voltage of 0·4 volts would result from conductivities $\sigma_n = 50\,(\Omega\,\text{cm})^{-1}$ and $\sigma_p = 33\,(\Omega\,\text{cm})^{-1}$, using Shockley's (1949) calculations of junction width and the value 2×10^5 volt cm^{-1} obtained by McAfee *et al.* (1951) for the field required to produce breakdown at a junction. It is not easy to see why the rather particular values of σ_n and σ_p should always be obtained under a wide variety of forming treatments.

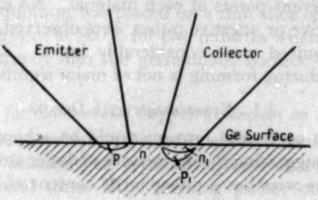


Figure 1.

Further, no satisfactory mechanism has been advanced to explain how p and n regions could be formed as a result of the power treatment applied to the collector. Under very heavy forming Valdes (1952) found a p-type region of diameter about 2×10^{-2} cm and depth about 10^{-3} cm to appear round a phosphor bronze point in n-type germanium, but after normal light forming the region affected was too small to be examined by Valdes' methods. Conditions in this case differ in that comparatively low temperatures over small areas, hardly extending beyond the collector point, obtain.

Experiments described in § 3, showing the possibility of forming with pure donor points, support these objections against the hook theory, since very artificial mechanisms must be proposed to account for any p region under the conditions of the experiments.

In view of the difficulties to which the preceding theories are subject, a further theoretical approach based on barrier modification by impurities introduced during forming has been considered. To determine the importance of these impurities, it appeared necessary to establish experimentally the significance of the composition of the point used during forming.

§ 3. EXPERIMENTAL RESULTS

In an effort to obtain more conclusive evidence from the somewhat varying data obtained from forming, experiments were performed with points which were either extremely pure donors or pure acceptors.

High purity aluminium (99·99%) and antimony (99·92%), both supplied by Johnson, Matthey and Co. were further purified by the zone melting technique, and point contacts prepared. The material of the points was expected to be as pure, or purer, than the original material.

Experiments were performed with these points as collectors subjected to forming treatments on two ground polished and etched single crystal n-type germanium surfaces. Surface A was of 5 Ω cm 100 μsec material etched with CP 4 etch and surface B was of 1½ Ω cm 50 μsec material etched with hydrogen peroxide (100 vols). Thorough and repeated etching was used.

Forming pulses of either polarity were supplied by discharging condensers through a variable resistance across the point contact and base. The results were displayed on an oscillograph such that α, R_{22} etc. could be measured immediately after a forming pulse.

Tungsten or phosphor bronze emitters were used in all cases at spacings of a few mils. In all cases a large variety of surface positions and pulses were used and a number of different points of each material. No significant differences in the results from positive or negative pulses were observed, though in general the positive voltages required were considerably lower. This suggests that the electric field present during forming is not of major significance.

3.1. *Experiments with Donors*

It was found that successful forming could be achieved with antimony but not with aluminium, suggesting the importance of donor atoms. Since commonly used phosphor bronze wire has a phosphorus content of 0·05% to 4% with tin from 1% to 7% and the balance copper, experiments were also performed with points made from commercial copper (containing traces of phosphorus), unpurified phosphor copper containing 15 wt% phosphorus, and unpurified phosphor tin of 5 wt% phosphorus.

The values of α (with I_e of order 1 mA) generally obtained at not too low R_{22} and corresponding values of R_{22} are shown in the table.

	Surface A		Surface B	
	α	R_{22}(kΩ)	α	R_{22}(kΩ)
Phosphor bronze	4–5	30	2–3	30
Antimony	2–3	20	2–3	20
Phosphor copper	4–5	20	2–3	20
Phosphor tin	4–5	20	2–3	20

With these materials R_{22} generally showed a progressive decrease with forming though transistor action remained down to values of R_{22} of a few thousand ohms. Further pulsing produced lower resistance ohmic contact with no hole collecting properties. The forward resistance remained low and similar to an unformed point in all cases.

In one case with the antimony collector an α of approximately 20 was observed, which was maintained in the presence of oscillations, up to a spacing of 20 mils before decreasing rapidly with further removal of the emitter. No satisfactory explanation of this phenomenon was adduced, but it is of interest to note that anomalous results such as this have been reported with phosphor bronze points.

Phosphor copper and phosphor tin gave similar results. On surface A in a few cases values of α of 8 to 10 with R_{22} in the range 10 to 15 kilohms were obtained. On surface B slightly lower values of α, 2 to 3, were obtained, under lesser forming pulses. On occasions however, values of α of 6 to 7 were obtained with the phosphor copper and in one case with the phosphor tin an α of 20, with oscillations, was observed. No such results were obtainable on this surface with phosphor bronze. The results tend to suggest that higher phosphorus content makes possible higher values of α in a greater percentage of cases.

3.2. *Results for Aluminium Formed Collectors*

Very heavy forming was necessary to reduce R_{22}. On surface A in a number of spots the discharge of an 8 μF condenser charged to ± 300 volts only reduced R_{22} to about 60 kΩ, from its initial value of several hundred kilohms. On surface B pulses of slightly less energy were required to reduce R_{22}. Here ohmic contact resulted after a pit was blown in the germanium surface with pulses much higher than those required to cause melting using phosphor bronze points.

It is not expected that the oxides film on the aluminium prevented its diffusion. It is noted that aluminium wires are formed on to silicon surfaces to give junction diodes. A pellet of aluminium was placed on a thin slice of germanium and the assembly heated in hydrogen at 500°C for two minutes when it was observed that the aluminium had diffused into the germanium surface forming a rectifying junction.

3.3. *Results for Commercial Copper Collectors on Surface A*

Varying results were obtained. On some spots, forming yielded values of α of 1 to 2. On other occasions all types of forming could not yield enhanced transistor action. Several times large forming pulses caused melting of the germanium accompanied by a flash and pits in the surface. The back resistance however remained high, values of 40 kΩ to 5 kΩ being variously observed. With all other metals used, flashing and pitting were invariably accompanied by the appearance of ohmic characteristics. Further, with the copper, there was still some transistor action after the melting had occurred.

We suggest that occasional very slight donor (phosphorus) impurities in the point were responsible for enhanced transistor action, and similarly, absence of such concentrations in the point of contact caused the frequently observed failure to show α enhancement on forming.

3.4. *Results for Gold Points on Surface A*

Using points of 99·99% pure Johnson Matthey gold, forming was successfully achieved with pulses rather higher than those used for phosphor bronze. The values of α were somewhat lower, a typical value being 2 for R_{22} about 10 kΩ.

Dunlap (1953) has found that gold introduces two high lying acceptor levels (0·15 ev and 0·5 ev) above the filled band in germanium, though Taft and Horn (1954) have found from resistivity temperature data of gold doped silicon that gold produces a donor level 0·33 ev above the full band in silicon. It appears that gold does not act in the same way as a simple Group III acceptor impurity in germanium. This would be particularly so in the surface region where the Fermi level is nearer the full band and well below the higher of the two energy levels introduced by the gold.

3.5. *Forming Breakdown with Indium p–n Junctions*

The explanation advanced for breakdown observed when point contacts on n-type germanium are subjected to heavy forming is based on the filling of surface acceptor levels by donors from the metal, resulting in progressive barrier lowering. Breakdown was however also observed with the pure acceptor aluminium points after extremely heavy forming, hence further experimental evidence on this phenomenon was sought by subjecting diffused indium p–n junctions to forming pulses.

After a certain pulse level the back resistance was observed to decrease markedly, and further pulsing resulted in breakdown with ohmic characteristics. This was observed to be due to ohmic contact being established at a few spots where melting occurred, on the indium pellet periphery. Further pulsing caused the indium at these spots to dissolve more germanium, leaving small holes in the 0·004 in. thick slab. The properties of the junction are apparently degraded by pulsing.

3.6. *Discussion of Results*

The experiments show that forming can be achieved using points containing donor impurities alone. The importance of donors has been previously suggested by Pfann (1951). The presence of Group III acceptor atoms does not appear to be required.

It is noted that Granville *et al.* (1955) have found experimentally that donor impurities are the significant ones for formed collectors on n-type silicon.

§ 4. THEORY

Formed point contacts share with unformed ones two important general characteristics, low forward resistance and characteristic rise in α at low I_e. This suggests that a basically different structure may not result from the forming process which is often very light.

In considering a diffusion model it is noted that Aarons *et al.* (1954) have found that atoms from gold points much larger than those used for transistors had passed into a germanium surface during forming. A technique for arcing antimony and aluminium into p- and n-type silicon crystal surfaces has been developed by Jacobs *et al.* (1954). The experiments of Granville *et al.* (1955) provide support for the assumption that impurities from the collector diffuse into the surface during forming.

If ϕ_0 is the collector barrier height and p_0 the excess hole concentration in the collector barrier region due to the emitter, we calculate $d\phi_0/dp_0$ in a manner similar to Hogarth's (1954) method for an unformed transistor, but with the addition of terms due to infused donor and acceptor concentrations.

We consider in n-type germanium of not too low resistivity a density n_s of ionized surface acceptor levels producing a rise ϕ_0 in electron energy at the surface $x=0$. Poisson's equation becomes

$$\frac{d^2\phi}{dx^2}=\omega\{(N_d-N_a)+p\exp(\phi/\mathbf{k}T)-n\exp(-\phi/\mathbf{k}T)+f_d-f_a\} \quad \ldots\ldots(1)$$

where $\omega=4\pi\mathbf{e}^2/\epsilon$, ϕ is electron energy measured from bottom of conduction band in the interior, $\mathbf{e}$ is electronic charge, ϵ dielectric constant of germanium, N_d and N_a donor and acceptor density normally present, and f_d and f_a infused donor and acceptor density respectively.

$$p=p_n+p_0+p_1 \quad \ldots\ldots(2)$$

where p_n is hole concentration before forming, p_1 hole concentration introduced by ionized infused acceptors and p_0 the hole concentration due to injection, assumed small.

$$n=n_n+n_1 \quad \ldots\ldots(3)$$

where n_n and n_1 are the electron concentrations before forming and introduced by the ionized infused donors respectively. We neglect the electrons which accompany the injected hole density p_0.

Equating surface and internal charges,

$$n_s = \int_0^{x_1} \{N_0 + p\exp(\phi/\boldsymbol{k}T) - n\exp(-\phi/\boldsymbol{k}T) + f_d - f_a\}\,dx \quad \ldots\ldots(4)$$

approximately if p_0 is small. We put $N_0 = N_d - N_a$ and x_1 is the value of x where ϕ equals zero and remains equal to zero for $x > x_1$. Then using the boundary conditions

$$\phi = \phi_0 \text{ at } x = 0, \qquad \frac{d\phi}{dx} = 0 \text{ at } \phi = 0 \quad \ldots\ldots(5)$$

we obtain

$$n_s = -\frac{1}{\omega}\left(\frac{d\phi}{dx}\right)_{x=0}. \quad \ldots\ldots(6)$$

It has been assumed that the forming has not destroyed the surface states entirely, though n_s after forming is probably less than before forming, making for a lower value of ϕ_0 and a reduced back resistance for a point contact. Similarly the value of x_1 would be different after forming.

The term $(d\phi/dx)_{x=0}$ in equation (6) is obtained by integrating equation (1). This is difficult to do exactly. We use as an approximate solution obeying the boundary conditions (5)

$$\frac{d\phi}{dx} = -\frac{\sqrt{(2\omega)}}{r}\{N_0\phi + \boldsymbol{k}Tp[\exp\phi/\boldsymbol{k}T - 1] + \boldsymbol{k}T(n_n + n_1)[\exp(-\phi/\boldsymbol{k}T - 1)]\}^{1/2} - \omega\int_x^{x_1}(f_d - f_a)\,dx \quad \ldots\ldots(7)$$

which is shown in the Appendix to be a fair approximation near $x = 0$, r being an integer between 1 and about 10.

With this approximation we find that the rate of change of collector barrier height with hole concentration due to the emitter,

$$\frac{d\phi_0}{dp_0} = -\boldsymbol{k}T\exp(\phi_0/\boldsymbol{k}T)\Big/\left[(N_0 + n_n + n_1) + \left\{\frac{\omega r^2}{2\boldsymbol{k}T}(n_s - F)^2 - \frac{N_0\phi_0}{\boldsymbol{k}T}\right\}\right] \quad \ldots\ldots(8)$$

where

$$F = \int_0^{x_1}(f_d - f_a)\,dx.$$

The expression is obtained by substituting equation (7) into (6) and extracting p_0 explicitly, using the approximation $\exp(\phi_0/\boldsymbol{k}T) \gg 1$. It may be noticed that terms containing $p_n + p_1$ do not appear in equation (8).

For the unformed case, $F = n_1 = 0$, $r = 1$, and

$$\frac{d\phi_0}{dp_0} = -\boldsymbol{k}T\exp(\phi_0/\boldsymbol{k}T)\Big/\left[(N_0 + n_n) + \left\{\frac{\omega}{2\boldsymbol{k}T}n_s^2 - \frac{N_0\phi_0}{\boldsymbol{k}T}\right\}\right] \quad \ldots\ldots(9)$$

where n_s and ϕ_0 have their pre-forming values. Equation (9) is similar to an expression which can be derived from equations given by Hogarth (1954) who however neglected the electron contribution n_n. The term $(2\pi e^2 n_s^2/\epsilon\boldsymbol{k}T)$ in equation (9) differs from the equivalent term given by Hogarth by a factor 4π, but numerical calculation confirms that this factor must be included.

§ 5. CALCULATION OF α

$$\alpha \equiv \left(\frac{dI_c}{dI_e}\right)_{V_c\,\mathrm{const}} = \frac{dI_c}{d\phi_0}\frac{d\phi_0}{dp_0}\frac{dp_0}{dI_e}. \qquad \ldots\ldots(10)$$

and $I_c = I_{cn} + I_{cp}$ where I_{cp} is the hole current to the collector and I_{cn} is equal to $B\exp(-\phi_0/\boldsymbol{k}T)$ (as approximately found by Gunn (1954), Simpson and Armstrong (1953) and earlier theory) B being a constant. The term dI_{cp}/dI_e cannot exceed 1 and is known experimentally to be small in the unformed case. We may assume, following Hogarth, that dI_{cn}/dI_e is the major term. Then

$$\alpha = B\frac{dp_0}{dI_e}\left(-\frac{1}{\boldsymbol{k}T}\exp\left(-\frac{\phi_0}{\boldsymbol{k}T}\right)\frac{d\phi_0}{dp_0}\right), \qquad \ldots\ldots(11)$$

and after forming,

$$\alpha = B\frac{dp_0}{dI_e}\Big/\left[(N_0+n_n+n_1)+\left\{\frac{\omega r^2}{2\boldsymbol{k}T}(n_s-F)^2-\frac{N_0\phi_0}{\boldsymbol{k}T}\right\}\right]. \qquad \ldots\ldots(12)$$

Hence α may be increased greatly by forming if F is positive and comparable with n_s, provided this is the dominant term. Numerical calculation (see Appendix) shows that this is so.

Since
$$F = \int_0^{x_1}(f_d - f_a)\,dx$$
the result immediately follows that donors alone can cause forming improvements whereas acceptors alone cannot, as found experimentally. We can define the forming improvement factor

$$\xi = \frac{N_0+n_n+(1/\boldsymbol{k}T)(2\pi \boldsymbol{e}^2 n_s{}^2/\epsilon - N_0\phi_0)}{N_0+n_n+n_1} \qquad \ldots\ldots(13)$$

as the ratio of the maximum value of α after forming to that before forming, where n_s and ϕ_0 have their pre-forming values. In order to estimate ξ we must assign values to n_s and ϕ_0. No accurate values for these quantities appear to have been obtained, but several estimates based on indirect methods have been published.

For materials of resistivity $2{\cdot}2\,\Omega$cm and $7\,\Omega$cm Hogarth (1953) has rather indirectly estimated n_s as about $11\times10^{10}\,\mathrm{cm}^{-2}$ and $4{\cdot}3\times10^{10}\,\mathrm{cm}^{-2}$ respectively with corresponding barrier heights in the range 0·25 to 0·27 ev and 0·166 to about 0·20 ev. Various other estimates of these orders have been given. In view of the uncertainties in the values of n_s and ϕ_0, we have calculated $d\phi_0/dp_0$ for an unformed transistor, and ξ for a range of values suggested by the above estimates. Results appear in figure 2 and 3.

It is apparent that values of ξ, especially at higher values of n_s, are adequate to account for the increased values of α observed after forming. Since the discussion refers to low emitter currents, a typical case might be an increase in α from 1 to 8 after forming. Values of ξ in excess of this are readily possible. Most forming of course does not obtain the optimum improvement which is possible.

If as an approximation we put $dp_0/dI_e = p_0/I_e$ in equation (11) we obtain

$$\alpha = \left(\frac{I_c}{I_e}p_0\right)\left(-\frac{1}{\boldsymbol{k}T}\frac{d\phi_0}{dp_0}\right). \qquad \ldots\ldots(14)$$

If we assume $10\,\mu$A as a reasonable value for I_c and I_e and $10\times10^{-14}\,\mathrm{cm}^3$ for $-(\boldsymbol{k}T)^{-1}d\phi_0/dp_0$ from figure 2 we find p_0 to be about $10^{12}\,\mathrm{cm}^{-3}$, and hence a value of 0·1 for α for an unformed transistor with 7 Ωcm germanium. The choice of different possible values of p_0, I_e and $d\phi_0/dp_0$ would still yield similar values for α. The result is in agreement with experiment at low emitter currents to which the calculation refers.

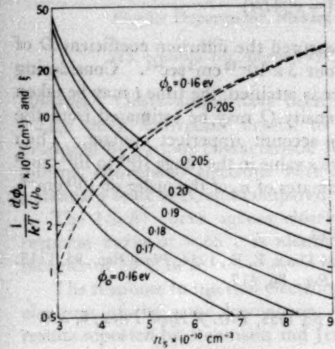


Figure 2. 7Ωcm Ge.

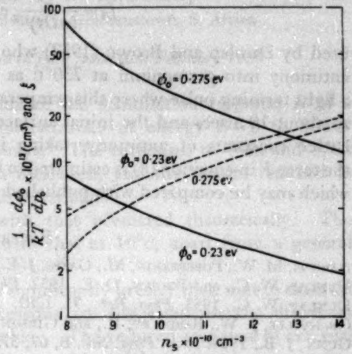


Figure 3. 2·2Ωcm Ge.

§ 6. CONCLUSION

The results show that current gain in transistors may be accounted for by the lowering of the collector barrier by injected holes. The effect is enhanced by the presence of infused donors in the region of the collector barrier after forming. Numerical calculations within the uncertainty of the values of surface state density and barrier height indicate that the values of α observed before and after forming may be accounted for on this mechanism.

ACKNOWLEDGMENTS

The benefit of discussions with Dr. L. W. Davies and members of the Laboratory is gratefully acknowledged. The author also wishes to thank Dr. E. W. J. Mitchell for helpful comments.

The work described in this paper was carried out as part of the programme of the Division of Radiophysics of the Commonwealth Scientific and Industrial Research Organization, Australia.

APPENDIX

Equation (7), the approximate solution to equation (10), is obtained by considering the exact solution (A) to an equation (B) like equation (1) but with the terms containing ϕ multiplied by a factor β which is slowly varying with respect to x at small x. Numerical calculation shows that the terms containing x are dominant in equations (1), (A) and (B). By considering the exact solution of equation (B) both for $\beta=0$ and for β having its calculated value of about 10 we can show that equation (7) is an approximation which is expected to be considerably closer than 10%.

For the case of forming with an antimony collector f_a and F may be calculated from the diffusion equation

$$f_a = \frac{Q}{(\pi Dt)^{1/2}} \exp(-x^2/4Dt)$$

used by Dunlap and Brown (1952) who measured the diffusion coefficient D of antimony into germanium at 730°c as about $3 \times 10^{-12}\,\text{cm}^2\,\text{sec}^{-1}$. Considering a light forming pulse where this temperature is attained, the time t may be taken as about 10 msec, and the initial surface density Q may be estimated from the lattice constants of antimony, taking into account imperfect contact. Then the term F in equation (8) is estimated to have a value in the range 10^{12} to $10^{13}\,\text{cm}^{-2}$ which may be compared with published estimates of n_s of the order of $10^{12}\,\text{cm}^{-2}$.

References

Aarons, M. W., Pobereskin, M., Gates, J. E., and Dale, E. B., 1954, *Phys. Rev.*, **95,** 1345.
Dunlap, W. C., and Brown, D. E., 1952, *Phys. Rev.*, **86,** 417.
Dunlap, W. C., 1953, *Phys. Rev.*, **91,** 1282.
Granville, J. W., Bardsley, R., and Gibson, A. F., 1955, *Brit. J. Appl. Phys.*, **6,** 206.
Gunn, J. B., 1954, *Proc. Phys. Soc.* B, **67,** 575.
Hogarth, C. A., 1953, *Proc. Phys. Soc.* B, **66,** 845; 1954, *Ibid.*, **67,** 636.
Jacobs, H., Brand, F. A., Matthei, W., and Ramsa, A. P., 1954, *J. Appl. Phys.*, **25,** 1406.
McAfee, K. B., Ryder, E. J., Shockley, W., and Sparks, M., 1951, *Phys. Rev.*, **83,** 650.
Pfann, W. G., 1951, *Phys. Rev.*, **81,** 882.
Seitz, F., 1951, *Phase Transformation in Solids* (New York : John Wiley). p, 77.
Shockley, W., 1949, *Bell Syst. Tech. J.*, **28,** 435; 1950, *Phys. Rev.*, **78,** 294.
Shockley, W., Sparks, M., and Teal, G. K., 1951, *Phys. Rev.*, **83,** 151.
Simpson, J. H., and Armstrong, H. L., 1953, *J. Appl. Phys.*, **24,** 25.
Sittner, W. R., 1952, *Proc. Instn Radio Engrs, N.Y.*, **40,** 448.
Taft, E. A., and Horn, F. H., 1954, *Phys. Rev.*, **93,** 64.
Valdes, L. B., 1952, *Proc. Instn Radio Engrs, N.Y.*, **40,** 445.

Scintillation Response of Anthracene to 6–30 kev Photoelectrons

By J. B. Birks† and F. D. Brooks‡
Physics Department, Rhodes University, Grahamstown, S. Africa

MS. received 3rd *August* 1955, *and in final form* 16*th February* 1956

Abstract. Observations have been made of the scintillation response S of a 0·2 cm thick anthracene crystal to photoelectrons of energy $E = 6$ to 30 kev produced by various characteristic K x-rays. The responses to ^{114}In and ^{137}Cs internal conversion electrons were also measured, and the results reduced to a standard scale of S, and compared with other data.

The (S, E) curve agrees closely with that predicted theoretically. The response curve at $-65°$c is identical with that at $16°$c, apart from a general increase of 50% in S.

The response to internal electrons is significantly greater than that to external electrons of the same low energy. Reasons are suggested for the divergent results reported by Robinson and Jentschke. The reduced response to external electrons is attributed to the escape and/or quenching of the fluorescence excitation energy at the crystal surface. These and other surface phenomena affecting the response of organic scintillators to short-range external particles are discussed.

§ 1. Introduction

When ionizing radiations impinge on anthracene or other fluorescent organic crystals, part of their incident energy is converted into light, which is emitted as normal fluorescence from the crystal. The scintillation response S depends on the nature and energy E of the incident ionizing particle, of range r within the crystal. Birks (1951) has shown that the specific fluorescence dS/dr may be expressed as a function of the specific energy loss dE/dr, and appears to be independent of the nature of the particle. The theoretical relation

$$\frac{dS}{dr} = \frac{A\,dE/dr}{1 + kB\,dE/dr} \qquad \ldots\ldots(1)$$

has been obtained by considering the competitive processes of fluorescence emission and of primary quenching by molecules damaged or ionized by the incident particle.

An alternative expression has been derived by Wright (1953), who attributes the primary quenching to the combined effect of unimolecular and bimolecular processes in the primary excitation column. This leads to a relation

$$dS/dr = C \ln(1 + D\,dE/dr) \qquad \ldots\ldots(2)$$

which approximates to (1), except at high dE/dr. Other relationships between dS/dr and dE/dr have been proposed by Chou (1952) and Black (1953).

---

† Now at British Dielectric Research Ltd., London, W.12.

‡ Now at the Atomic Energy Research Establishment, Harwell, Berks.

These expressions account satisfactorily for the scintillation response of organic crystals (and also solutions) to all types and energies of incident ionizing particles, except those of short range. For α-particles of r less than 8 mm air equivalent, the value dS/dr is, however, less than that given by (1) or (2) (Birks 1950, King and Birks 1952). A similar effect is observed with short-range electrons and protons (Taylor *et al.* 1951). Since both effects are associated with short-range particles it is suggested that they are due to surface phenomena, probably either surface escape of excitons or radiations (Birks 1952, 1953) or quenching of fluorescence excitation energy by a non-fluorescent impurity in the surface layer (King 1952). A theory of surface escape has been formulated by Birks (1952, 1953). It involves multiplying the right-hand side of (1) or (2) by a 'surface escape function'

$$\phi = 1 - \tfrac{1}{2}[\exp(-r/a_0) - (r/a_0)\mathrm{Ei}(r/a_0)] \qquad \ldots\ldots(3)$$

where $\mathrm{Ei}(r/a_0)$ is the exponential integral, and a_0 the mean free path of the escaping photons. ϕ increases from 0·5 at $r = 0$, to approximately 1·0 for $r > 2a_0$. Recent observations by Wright (1955) indicate that quenching impurities are formed on the surface of anthracene crystals exposed to the atmosphere, thus causing surface quenching.

If surface effects are operative, the scintillation response of anthracene to photoelectrons produced *within* the crystal by x-rays should differ from that excited by external electrons of the same energy, since in the former case surface effects are eliminated (Birks 1952). A detailed study has therefore been made of the scintillation response of anthracene to photoelectrons of 6–30 kev energy for comparison with the external electron data of Taylor *et al.* (1951). A brief account of the preliminary results, which showed the predicted difference in response, has been published previously (Birks and Brooks 1954). Robinson and Jentschke (1954) have subsequently reported the results of a similar investigation, but showing no difference in the response to internal and external electrons. The present paper gives a fuller account of our original and later studies. These substantiate the differences in response originally reported. Similar results have recently been obtained by Fowler and Roos (1955).

§ 2. EXPERIMENTAL METHOD

2.1. *Apparatus*

A schematic diagram of the apparatus is shown in figure 1. The purpose of the experiment was to measure the scintillation response of the anthracene crystal D (0·2 cm × 0·5 cm × 0·5 cm) to photoelectrons released within it by

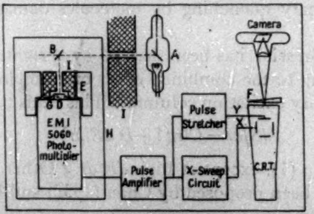


Figure 1. Schematic diagram of apparatus.

A, x-ray tube ; B, scattering foil ; D, anthracene crystal ; E, cooling jacket ; F, wedge filter ; G, aluminium reflector ; H, light-proof box ; I, lead shielding.

secondary characteristic fluorescence K x-rays of different elements. Foils of these elements were placed at B, directly above the crystal and in the path of a collimated beam of 'white' x-radiation, AB. Pure metal foils ($4\,\mathrm{cm}^2 \times 0{\cdot}01$ cm thick) of Fe, Cu, Mo, Ag and Sn were used. Powdered layers of As, I and SrO, mounted in flat black paper packets, served as 'foils' of their respective elements. With these elements an x-ray energy range of 6 kev to 32 kev was obtained. A mechanical arrangement was used, enabling the foils to be changed or radioactive sources to be substituted, without exposing the photomultiplier to external light. During the investigation, various x-ray tubes were used as sources of 'white' radiation, and the results were found to be independent of the tube used.

Other low-energy x-ray sources were also employed, as follows: (*a*) In K x-rays emitted following internal conversion in ^{114}In. Internal conversion electrons from the source were eliminated by a Perspex absorber. (*b*) Mn K x-rays emitted following K-capture in ^{55}Fe.

The scintillations were observed with an E.M.I. type 5060 photomultiplier, selected for its low dark noise and high sensitivity. The crystal D and the photocathode could be cooled by refrigerants placed in the cooling jacket E. By this means the dark noise could be virtually eliminated, and the temperature dependence of S could be studied.

Voltage pulses from the photomultiplier were fed to an A.E.R.E. Model 1049 A linear pulse amplifier. The pulse amplitude distribution was measured with a multi-channel photographic wedge pulse analyser (Maeder 1947), based on the design of Bernstein *et al.* (1953). This type of pulse analyser is particularly suitable for studying pulse sequences initiated by sources of fluctuating intensity. Provided the counting rate is not excessive, practically every pulse is counted. The only limitations on the amplitude resolution are the width of the cathode-ray tube trace and the circuit stability during the relatively short period of measurement. As described below, the characteristic response of the photographic film can be used to obtain a sharply defined pulse amplitude distribution.

2.2. *X-ray Absorption in Anthracene*

Internal photoelectrons of known energy are efficiently produced in anthracene by x-rays of energies below 20 kev since the cross section of the photoelectric absorption process in carbon (see Heitler 1949 for data) is relatively high. At x-ray energies greater than about 22 kev however, Compton scattering becomes the main absorption process, and the photoelectric absorption declines rapidly. This places an upper limit on the photoelectron energy which can be resolved against the 'Compton background', and against the background due to scattered 'white' x-rays which, though continuous and of low intensity, are efficiently absorbed at low energies. The measurements indicated 30 kev as the upper energy limit at which a photoelectric peak could be clearly resolved with the arrangement used.

2.3. *Photoelectron Energy*

The K_α and K_β components in the characteristic fluorescence were not resolved. A weighted mean photoelectron energy for each foil used was calculated from the following data: (*a*) the K_α and K_β x-ray energies (Hodgman 1954), (*b*) the $K_\alpha : K_\beta$ intensity ratio, which varies from 5 : 1 to 3 : 1 for the elements used (Compton and Allison 1946), (*c*) the photoelectric absorption coefficient. The values obtained are listed in the table.

Atomic number	25	26	29	33	38	42	47	49	50	53
Element	Mn	Fe	Cu	As	Sr	Mo	Ag	In	Sn	I
K_α energy (kev)	5·9	6·4	8·1	10·5	14·2	17·5	22·2	24·2	25·2	28·6
K_β energy (kev)	6·3	7·1	9·0	11·8	16·0	19·9	25·3	27·7	28·9	32·8
Weighted mean photoelectron energy (kev)	5·9	6·5	8·2	10·7	14·5	17·9	22·7	24·8	25·8	29·4

2.4. *Calibration and Test*

The scintillation response of the anthracene crystal to external electrons of known high energy was also measured. This enabled the photoelectron measurements to be normalized to the same scale of S as that used by other observers. Internal conversion electrons from ^{114}In (186 and 162 kev) and ^{137}Cs (624 kev) were used for this purpose. The two electron groups from ^{114}In were not clearly resolved, but the mean pulse amplitude was observed. The ratio of the observed response to ^{114}In and ^{137}Cs electrons was in agreement with the measurements of Taylor *et al.* (1951), Hopkins (1950, 1951) and other observers.

The comparative measurement of S over the interval from 6 kev to 624 kev electron energy involved accurate calibration of the pulse amplifier gain and the photomultiplier gain over a range of more than 100 : 1. This calibration was performed using standard scintillation sources (i.e. mono-energetic α-particles, or ^{137}Cs internal conversion electrons, on anthracene) and comparing the mean pulse amplitudes at different gain settings. For this phase of the work, in which sources of constant intensity were used, the pulse amplitude distributions were measured with a discriminator and scaler. The gain calibration was shown to be consistent by numerous cross-checks, and was found to remain stable over the whole period of the measurements. Various other instrumental tests and calibrations were also used to ensure optimum and reliable performance.

2.5. *Pulse Amplitude Distribution*

Two typical photographs of the pulse amplitude distribution records, using As K x-rays and Sn K x-rays, are shown in figure 2 (*a*) and (*b*) respectively. Figure 2 (*b*) shows the increase in low amplitude background at the higher x-ray energies, referred to in § 2.2. The records were obtained on 35 mm Kodak ‘Microfile’ film, using a Zeiss Contax Camera. The enlargements were made on to high-contrast reflex document copying paper, and they are inverted and rotated through 90° relative to the presentation on the cathode-ray tube of the wedge analyser.

The use of high-contrast photographic materials gives a sharp transition from black to white, which serves as an iso-density contour or edge, from which the pulse amplitude distribution can be deduced (Maeder 1947). Since linear amplification is used, the pulse amplitude is proportional to the ordinate of the edge. N, the number of pulses per unit pulse height, is related to the abscissa x of the edge by

$$x = R \ln (N/N') \qquad \ldots\ldots (4)$$

where R and N' are constants, determined by the wedge density and the camera lens aperture respectively (Brooks 1955). Calibrations in terms of (4) were undertaken, and the abscissae in figure 2 are shown in units of N. For the present studies an accurate determination of N was not required.

2.6. *Scintillation Response*

The peak, or most probable, value of the pulse amplitude was obtained directly from the photographic records. At low particle energies this peak value is not exactly proportional to the scintillation response S, due to the asymmetry of the pulse amplitude distribution (Wright 1954 a). The peak pulse amplitude data were therefore corrected to give the true values of S, by the procedure described by Wright (1954 b).

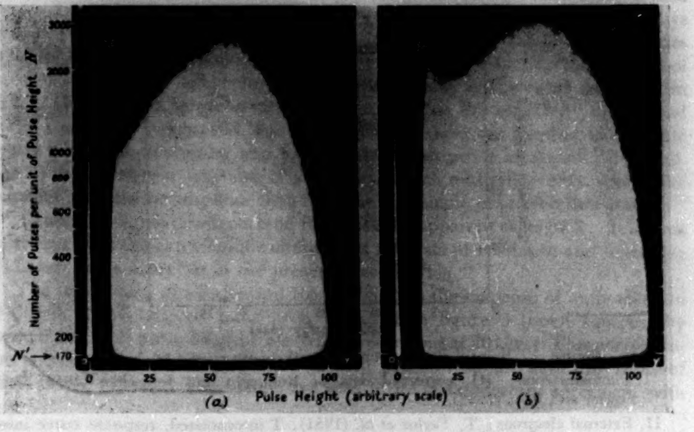


Figure 2. Pulse amplitude distribution photographs.
(*a*) As K x-rays (mean photoelectron energy = 10·7 kev)
(*b*) Sn K x-rays (mean photoelectron energy = 25·8 kev)

§ 3. EXPERIMENTAL RESULTS

The observed values of the scintillation response S of the anthracene crystal to photoelectrons of energy E are plotted in figure 3 (extended points). The scale of S is the same as used by Taylor *et al.* (1951) in their measurements of the response to external electrons ($S = 6240$ units at $E = 624$ kev).

For each photoelectron energy E, three or four separate measurements were made at room temperature (∼16°c). In addition one measurement was made at a temperature of $-65°$c ($\pm 2°$c) for each energy, except for I, ^{114}In and ^{55}Fe. At $-65°$c, the scintillation response to 624 kev electrons increased by a factor of 1·50 relative to the response at room temperature. The response to photoelectrons from 6–30 kev energy increased by the same factor. Hence the low temperature photoelectron response data were normalized in terms of the 624 kev electron response, and included with the room temperature data in obtaining the mean values of S plotted in figure 3.

The error limits shown in figure 3 include the probable errors from the pulse amplitude measurements, from the instrumental calibrations, and from the normalization of the data.

§ 4. DISCUSSION

4.1. *Present Results*

The data shown in figure 3 differ slightly from the preliminary results reported previously (Birks and Brooks 1954), and to which the statistical correction (§ 2.6) was applied by Wright (1954 b). The difference is due to the inclusion of additional observations. It will be noted that the corrected response S is not proportional to the photoelectron energy E, although it is approximately linear over the energy range considered.

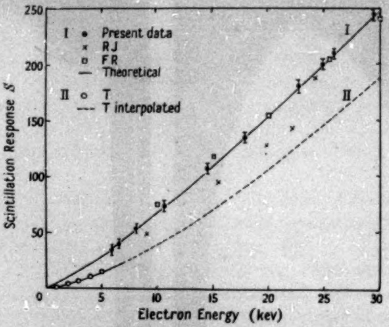


Figure 3. Scintillation response of anthracene crystal to electrons.

I, Photoelectrons: RJ, corrected data of Robinson and Jentschke (1954); FR Fowler and Roos (1955).

II, External electrons: T, Taylor *et al.* (1951); T interpolated, response curve interpolated between 5 and 80 kev by Taylor *et al.* (1951).

The increase in the scintillation response of anthracene by a factor of 1·50 in cooling from 16°c to −65°c agrees with previous observations (Sangster and Irvine 1956, Liebson 1952). The fact that the shape of the relative response curve is unaffected by the temperature reduction is significant. It indicates that the increased efficiency at low temperatures is not due to a reduction in the primary quenching processes, responsible for the non-linear response.

4.2. *Comparison with other Results*

The experimental data of Taylor *et al.* (1951) on the response of anthracene to external electrons of 1–5 kev energy are plotted in figure 3 (open circles).

Robinson and Jentschke (1954) have reported observations on the response to photoelectrons, agreeing with the external electron response data of Taylor *et al.* They did not, however, correct for the asymmetry of their pulse-height distributions (§ 2.6). The corrected values, estimated from their published observations, are plotted in figure 3 (crosses); they are significantly higher than those of Taylor *et al.*†

---

† In addition there appears to be a discrepancy, with regard to the response to In K x-rays, between the text of the paper by Robinson and Jentschke and their figure 5. The plotted point $S=180$ units at $E=28$ kev would appear from the text to refer to the In K_α energy which is 24·2 kev. Assuming the latter to be correct, a corrected value of $S=190$, $E=24{\cdot}2$ kev is obtained and plotted, and it is in good agreement with the present results.

The remaining discrepancies between the present results and those of Robinson and Jentschke may be attributed to differences in experimental method and arrangement, which were notably as follows: (i) Robinson and Jentschke used a single-channel pulse analyser, involving an estimated measuring period of at least 3–4 hours. This may be compared with a period of less than 30 minutes with the multi-channel wedge analyser, which thus reduces errors due to instrumental instability. (ii) They needed to subtract background counts up to pulse amplitudes equivalent to $E\sim15$ kev. In the present measurements background noise was negligible even at $E\sim6$ kev. (iii) They used a larger anthracene crystal ($2\,\text{cm}\times2\,\text{cm}\times1\,\text{cm}$) and different x-ray sources, mainly excited by In K x-rays from ^{113}Sn, after the method described by Insch (1950). The various activities associated with ^{113}Sn (Hollander *et al.* 1953) probably produced a background of secondary radiations, analogous to the 'white' x-ray background described in § 2.2. The use of a larger crystal would accentuate the sensitivity to this and to the Compton background, and would also increase the number of multiple events (Compton scattering and photoelectric absorption in cascade) detected. Due to the non-linear response of anthracene, such multiple events would give a smaller scintillation response than that due to a single photoelectric event.

All such background effects tend to reduce the apparent value of S. It is thus consistent with the explanations offered that the data of Robinson and Jentschke fall below those observed in the present investigation.

The recent data of Fowler and Roos (1955) on the response of anthracene to 10–30 kev photoelectrons have been included in figure 3 (open squares) by normalizing their pulse-height scale to the present one at 20 kev. This normalization is consistent with a comparison of their proton response data with those of Taylor *et al.* (1951). The results are seen to be in satisfactory agreement with ours.

§ 5. THEORY

5.1. *Specific Fluorescence and Response to Photoelectrons*

The scintillation efficiency dS/dE has been calculated as a function of E from the observed (S, E) values (figure 3) for photoelectrons. The specific energy loss dE/dr as a function of E has been similarly obtained from the range–energy data of Curie (1935) for electrons in standard air (all range data used refer to *integrated* ranges in standard air). Hence the specific fluorescence dS/dr for photoelectrons from 6–30 kev energy has been determined as a function of dE/dr.

Similar calculations have been applied to the (S, E) values for protons observed by Taylor *et al.* (1951) and Franzen *et al.* (1950), using the proton range–energy data of Livingston and Bethe (1937), and $(dS/dr, dE/dr)$ has thus been obtained for protons.

It is found that the $(dS/dr, dE/dr)$ curves are identical for electrons and protons, within the experimental error and scope of the present and quoted measurements. The original hypothesis (Birks 1951) that dS/dr is a monotonic function of dE/dr, and independent of the nature of the ionizing particle (in the absence of surface effects) is thus verified for electrons and protons.

The experimental $(dS/dr, dE/dr)$ data are equally consistent with (1), (2) and the other formulae referred to in § 1. The constants in (1) and (2) have been evaluated from the experimental response and range data, and the theoretical response curve shown in figure 3 (solid line) has been obtained by graphical integration.

At higher energies the theoretical response is also in agreement with the external electron response data of Taylor *et al.* (1951), Hopkins (1950, 1951), Robinson and Jentschke (1954) and of the present investigation (§2.4). These show a linear response for anthracene to electrons with E greater than 120 kev, which extrapolates to intersect the E axis at $E = 20$–25 kev.

This feature of the experimental response curve is inherent in the theory. At high electron energies, where dE/dr is small, (1) or (2) may be expanded and integrated, giving

$$S = A\int_0^r (dE/dr)(1 - kB\,dE/dr)\,dr$$

$$= A(E - E') \qquad \ldots\ldots(5)$$

where

$$E' = kB\int_0^E (dE/dr)\,dE. \qquad \ldots\ldots(6)$$

E' is practically constant for $E > 200$ kev, since dE/dr is small and decreases with increasing E. A value of $E' = 24$ kev has been calculated from kB and the range–energy data, and this corresponds from (5) to the intercept on the E axis of the linear portion of the (S, E) curve.

Although the response curve (figure 3) is approximately linear for low energy electrons, the scintillation efficiency is less than at higher energies. At $E = 20$ kev, $dS/dE = 9{\cdot}0$ units of S/kev, compared with $dS/dE = 10{\cdot}4$ for $E > 120$ kev. This decrease is attributed primary quenching processes as described by (1) or (2), and it is distinct from the additional surface effect encountered with externally incident electrons.

5.2. *Surface Effect for External Electrons*

The experimental results (figure 3) show a significant difference in the scintillation responses to internal photoelectrons and to external electrons. It is clear that the scintillation efficiency of anthracene is reduced for electrons absorbed near the crystal surface. This surface effect might be caused by one or more of the following processes: (*a*) back-scattering of the primary electrons; (*b*) additional quenching of the fluorescence near the crystal surface, due to chemical deterioration or other modification of the surface layer; and (*c*) surface escape of photons, emitted during the scintillation process, which would otherwise be absorbed in the crystal, leading to further fluorescence emission, and thus contributing to the scintillation response.

The effect of back-scattering has been considered by Taylor *et al.* (1951) and Butt (1952, 1953). The former estimate that less than 10% of the incident 1–5 kev electrons are scattered in a backward direction. The latter concludes that for 4–8 kev incident electrons, at least 80% of the back-scattered electrons, which constitute approximately 10% of the incident beam, dissipate half their energy in the crystal. We have made similar calculations based on the recent data of Sternglass (1954) on the back-scattering of 1–10 kev electrons from carbon, and have obtained results in agreement with Butt. It is estimated that the *maximum* effect of back-scattering of the incident electrons will be to cause a 4% reduction in the scintillation response observed by Taylor *et al.* (1951). This is quite inadequate to account for the magnitude of the observed surface effect.

Wright (1955) has recently made measurements of the photo-fluorescence excitation spectra of anthracene crystals. These show a marked dependence of the flourescence efficiency on the depth of penetration of the exciting radiation, the efficiency being less when the radiation is absorbed close to the surface. The magnitude of the effect depends on the state of the crystal surface, and appears to be partly due to chemical or physical deterioration of the surface, as well as to the escape of normal fluorescence radiation.

The surface effect has previously been attributed to the surface escape of radiation as described by (3) (Birks 1952, 1953). A mean free path a_0 for the escaping radiations of 3 to 8μ was found to give a reasonably consistent description of the external electron data of Taylor *et al.* (1951) and the α-particle data of King and Birks (1952). Emission and absorption processes in cascade are known to occur in anthracene in the 4000–4500 Å wavelength region (Birks and Little 1953). It has also been suggested that similar processes may occur in the 1650–1800 Å wavelength region during the scintillation process (Birks 1953, 1954 a). Hence surface escape of the normal (longer wavelength) fluorescence is to be expected, while escape of the hypothetical primary (short wavelength) radiation is also possible. These alternatives cannot, however, be readily differentiated using the electron response data for anthracene, nor can the possible contribution of surface quenching be estimated without detailed knowledge of the crystal used by Taylor *et al.* (1951).

It is of interest to note that King and Birks (1952) have observed a surface effect of similar magnitude and a_0 in anthracene, *para*-terphenyl and *trans*-stilbene crystals. The overlap of the normal emission and absorption spectra in these last two materials is very much less than in anthracene (Birks and Wright 1954) and it is indeed negligible in *para*-terphenyl. Thus any surface effect associated with the escape of normal fluorescence should be much reduced in these materials, compared with anthracene, and be insignificant in the case of *para*-terphenyl.

A similar decrease in the scintillation efficiency of terphenyl solutions, when excited by short-range α-particles of r less than 10 mm (equivalent from (3) to $a_0 \sim 10\mu$) has been observed by Reynolds (1952). Such solutions are highly transparent to the normal terphenyl emission. This analogous effect in liquid scintillators, which has been previously attributed to primary photon escape (Birks 1953), might also be influenced by surface quenching due to absorbed oxygen (Pringle *et al.* 1953).

Without a fuller knowledge of the actual experimental surfaces used in the different studies, it is not possible to assess properly the relative importance of the alternative photon escape and surface quenching processes. The similar dependence of dS/dr on dE/dr, and the similar surface effect observed for each of the different organic crystalline and liquid scintillators studied, suggests, however, that common primary quenching and surface processes are operative.

§ 6. CONCLUSIONS

The present investigation has clarified several matters associated with the scintillation process in organic crystals. It has shown that:

(*a*) The scintillation response of anthracene to internal electrons can be adequately described theoretically, and that the variation of dS/dr with dE/dr for electrons and protons is independent of the nature of the particles.

(*b*) The relative scintillation response of anthracene to electrons of different energies is unaffected by a reduction in temperature from 16°C to -65°C, indicating that the 50% increase in scintillation efficiency is not due to a reduction in primary quenching processes in the excitation column.

(*c*) The scintillation response of anthracene to internal electrons is greater than that to external electrons of the same low energy. This is the result of a surface effect, probably due to the escape of photons and/or to impurity quenching of the excitation energy. The effect, which appears to be characteristic of the organic scintillators, will be encountered with any short-range externally incident particles.

The surface effects in anthracene and other organic scintillators merit further investigation. Further data are required on the scintillation response to external electrons of 5–80 kev energy, suitable for comparison with the theoretical relation (3). In future studies attention should be paid to the state of the crystal surface, to allow for any surface quenching effects.

ACKNOWLEDGMENTS

We wish to thank our colleague, Dr. G. T. Wright, for access to his unpublished results and for several interesting discussions. We are indebted to Capt. K. G. F. Collender of the Witwatersrand Native Labour Association for the loan of an x-ray tube, and to the South African Council for Scientific and Industrial Research for a research grant and for a bursary to one of us (F.D.B.).

REFERENCES

Bernstein, W., Chase, R. L., and Schardt, A. W., 1953, *Rev. Sci. Instrum.*, **24,** 437.

Birks, J. B., 1950, *Proc. Phys. Soc.* A, **63,** 1294; 1951 a, *Ibid.*, A, **64,** 874; 1951 b, *Phys. Rev.*, **84,** 364; 1952, *Ibid.*, **86,** 569; 1953, *Scintillation Counters* (London: Pergamon Press; New York: McGraw-Hill); 1954 a, *Phys. Rev.*, **94,** 1567; 1954 b, *Ibid.*, **95,** 277.

Birks, J. B., and Brooks, F. D., 1954, *Phys. Rev.*, **94,** 1800.

Birks, J. B., and Little, W. A., 1953, *Proc. Phys. Soc.* A, **66,** 921.

Birks, J. B., and Wright, G. T., 1954, *Proc. Phys. Soc.* B, **67,** 657.

Black, F. A., 1953, *Phil. Mag.*, **44,** 263.

Brooks, F. D., 1955, *M.Sc. Thesis*, Rhodes University.

Butt, D. K., 1952, *Physica*, **18,** 242; 1953, *Proc. Phys. Soc.* A, **66,** 940.

Chou, C. N., 1952, *Phys. Rev.*, **87,** 904.

Compton, A. H., and Allison, S. K., 1946, *X-rays in Theory and Experiment* (New York: van Nostrand), p. 641.

Curie, M., 1935, *Radioactivité* (Paris: Hermann et Cie).

Fowler, J. M., and Roos, C. E., 1955, *Phys. Rev.*, **98,** 996.

Franzen, W., Pelle, R. W., and Scherr, R., 1950, *Phys. Rev.*, **79,** 742.

Heitler, W., 1949, *The Quantum Theory of Radiation* (Oxford: Clarendon Press).

Hodgman, C. D. (Ed.), 1954, *Handbook of Chemistry and Physics* (Cleveland: Chemical Rubber Publishing Co.)

Hollander, J. M., Perlman, I., and Seaborg, G. T., 1953, *Rev. Mod. Phys.*, **25,** 469.

Hopkins, J. I., 1950, *Phys. Rev.*, **77,** 406; 1951, *Rev. Sci. Instrum.*, **22,** 29.

Insch, G. M., 1950, *Phil. Mag.*, **41,** 857.

King, J. W., 1952, *M.Sc. Thesis*, Rhodes University.

King, J. W. and Birks, J. B., 1952, *Phys. Rev.*, **86,** 568.

Liebson, S. H., 1952, *Nucleonics*, **10,** no. 7, 41.

Livingston, M. S., and Bethe, H. A., 1937, *Rev. Mod. Phys.*, **9,** 245.

Maeder, D., 1947, *Helv. Phys. Acta*, **20,** 139.

Pringle, R. W., Black, L. D., Funt, B. L., and Sobering, S., 1953, *Phys. Rev.*, **92,** 1582.

Reynolds, G. T., 1952, *Nucleonics*, **10,** no. 7, 46.

Robinson, W. H., and Jentschke, W., 1954, *Phys. Rev.*, **95,** 1412.

Sangster, R. C., and Irvine, J. W., 1956, *J. Chem. Phys.*, **24,** 670.

Sternglass, E. J., 1954, *Phys. Rev.*, **95,** 345.

Taylor, C. J., Jentschke, W. K., Remley, M. E., Eby, F. S., and Kruger, P. G., 1951, *Phys. Rev.*, **84,** 1034, and private communication.

Wright, G. T. 1953, *Phys. Rev.*, **91,** 1282; 1954 a, *J. Sci. Instrum.*, **31,** 377, 462; 1954 b, *Phys. Rev.*, **96,** 569; 1955 a, *Ibid.*, **100,** 587; 1955 b, *Proc. Phys. Soc.* B, **68,** 701.

On the Convection of Heat from a Wire moving through Air close to a Cooling Surface

BY THE LATE N. A. V. PIERCY[†], E. G. RICHARDSON[‡]
AND H. F. WINNY[§]

[†] Late Professor of Aeronautical Engineering in the University of London at Queen Mary College

[‡] Physics Department, King's College, Newcastle-upon-Tyne
[§] The Fairey Aviation Co. Ltd., Hayes, Middx.

MS. received 15*th November* 1955, *and in revised form* 16*th February* 1956

Abstract. Renewed interest in the measurement of fluid velocities in boundary layers by the hot-wire method has called attention to a difficulty that arises when the wire approaches a conducting body of lower temperature. The additional convection caused by the body may be large and is not easy to assess by experiment alone, while theoretical investigation is somewhat complicated. Some time ago, the present authors obtained an approximate solution of the cognate problem of a hot wire moving adjacently to a cooling flat plate through an inviscid fluid, assumed to be initially stationary, and found a measure of agreement with corresponding experiments which they conducted in air. It is anticipated that the results should provide reliable corrections in the practical problem when the wire is small compared with the thickness of the boundary layer and sufficiently clear of the surface of the body.

Section 1 of the present paper gives an account of the analysis, which is in the nature of a development of the work of King and of Piercy and Winny. Section 2 records observations of the cooling of a hot wire whirled through air round a large circular brass cylinder of normal temperature. The results of the experiments are compared with the theory by means of examples worked from the final equations.

§ 1. THEORETICAL

THE analytical problem is assumed to be two-dimensional, and the fluid to be inviscid, incompressible and of uniform thermometric conductivity. The fluid is assumed to have a uniform undisturbed velocity $-U$ parallel to an infinite plane boundary of infinite conductivity and heat capacity. In this z plane the axis of x is located on the boundary and the centre of the trace of the wire, of radius a, is at $x=0$, $y=b$.

The familiar solution of this potential motion is noted for future reference. In terms of the coordinate $t=\xi+i\eta$, defined by

$$z=\alpha\tan\tfrac{1}{2}t, \qquad \ldots\ldots(1)$$

$w=\phi+i\psi$ is given by

$$w=2U\alpha\sum_{1}^{\infty}(-)^s\exp(-s\eta_0)\frac{\sin st}{\sinh s\eta_0}-Uz. \qquad \ldots\ldots(2)$$

From (1)

$$x,y=\alpha(\sin\xi,\sinh\eta)/(\cos\xi+\cosh\eta) \qquad \ldots\ldots(3)$$

so that $\eta=0$ represents the x axis and on this streamline $\psi=0$. When $\eta=\eta_0$, (3) gives

$$x^2+(y-\alpha\coth\eta_0)^2=\alpha^2\operatorname{cosech}^2\eta_0,$$

while (2) gives

$$\psi=-\alpha U, \qquad \ldots\ldots(4)$$

so that the circle

$$a=\alpha\operatorname{cosech}\eta_0, \qquad b=\alpha\coth\eta_0 \qquad \ldots\ldots(5)$$

is this streamline. We require the variation with a and b of the maximum and minimum values of ϕ on the circle, denoted by $\pm\phi_0$, and this has been plotted in figure 1.

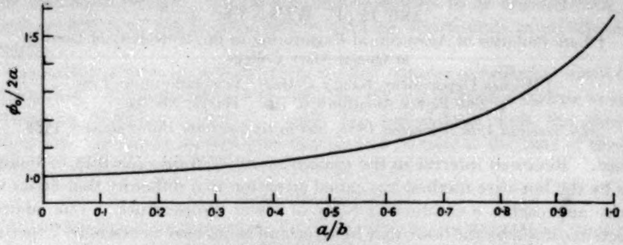


Figure 1. The function $\phi_0/2a$ in relation to a/b ; a radius of wire, b distance of axis from wall.

Denoting by k and c the constant conductivity and heat capacity of the fluid per unit volume, the equation for the temperature θ at any point in the fluid is

$$u\frac{\partial\theta}{\partial x}+v\frac{\partial\theta}{\partial y}=\frac{k}{c}\nabla^2\theta \qquad \ldots\ldots(6)$$

or, in the w plane,

$$\frac{\partial\theta}{\partial\phi}=\frac{k}{c}\left(\frac{\partial^2\theta}{\partial\phi^2}+\frac{\partial^2\theta}{\partial\psi^2}\right). \qquad \ldots\ldots(7)$$

In the w plane, $\psi=0$ represents the plane boundary, while the circle becomes both sides of the short length of the straight line $\psi=-\alpha U$ between $\pm\phi_0$ (figure 1). The thermal boundary conditions in this plane are specified by those in the z plane. It is readily verified that except at large Reynolds numbers, which are excluded, the temperature at the surface of a solid wire is closely uniform. Thus the temperature of the short plate in the w plane is assumed uniform, and denoted by θ_0. The temperature of the infinite plane boundary is assumed to be the same as that of the distant fluid and, without loss of generality, is represented as zero. This plane is replaced by the image in it of the short plate, and the temperature of this image is consistently denoted by $-\theta_0$. Hence the temperature of the fluid is $\pm\theta_0$ along $\psi=\mp\alpha U$ between the limits $\phi=\pm\phi_0$, and is zero along $\psi=0$.

Putting, as usual, $2n=c/k$ and $\theta=\theta_1 e^{n\phi}$, (7) becomes

$$-n^2\theta_1=\frac{\partial^2\theta_1}{\partial\phi^2}+\frac{\partial^2\theta_1}{\partial\psi^2}, \qquad \ldots\ldots(8)$$

which has the solution

$$\theta_1=C\mathrm{K}_0\{n(\phi^2+\psi^2)^{1/2}\}, \qquad \ldots\ldots(9)$$

where the Bessel function is defined generally by

$$\mathrm{K}_n(\xi)=\int_0^\infty \exp(\xi\cosh t)\cosh nt\,dt. \qquad \ldots\ldots(10)$$

Therefore,

$$\theta=Ce^{n\phi}\mathrm{K}_0\{n(\phi^2+\psi^2)^{1/2}\}. \qquad \ldots\ldots(11)$$

If Q is the quantity of heat issuing per second from unit length of a line source at the origin, $C=Q/2\pi k$, and a solution may be built up from a distribution of such sources. The present problem will yield to calculation of such distributions along the real plate and its image and, from symmetry, the two systems will differ only in sign. Let $\chi(\phi_1)$ be the heat flux from the real plate at $\phi=\phi_1$. The function is determined by the integral equation

$$2\pi k\theta_0=\int_{-\phi_0}^{\phi_0}\exp\{n(\phi-\phi_1)\}\{\mathrm{K}_0(n[\phi-\phi_1])-\mathrm{K}_0(n\{[\phi-\phi_1]^2+[2\alpha U]^2\}^{1/2})\}\chi(\phi_1)\,d\phi_1, \qquad \ldots\ldots(12)$$

in which it will be assumed that the maximum value of $\phi-\phi_1$ on the real plate or its image is small compared with $2\alpha U$, their distance apart.

Assuming for the solution the form

$$\chi(\phi_1)=A\chi_0(\phi_1)+B\chi_1(\phi_1), \qquad \ldots\ldots(13)$$

where A, B are constants and χ_0, χ_1 are defined by the integral equations

$$1=\int_{-\phi_0}^{\phi_0}\exp\{n(\phi-\phi_1)\}\mathrm{K}_0\,|n(\phi-\phi_1)|\,\chi_0(\phi_1)\,d\phi_1, \qquad \ldots\ldots(14)$$

$$1=\int_{-\phi_0}^{\phi_0}\exp(-n\phi_1)\mathrm{K}_0\,|n(\phi-\phi_1)|\,\chi_1(\phi_1)\,d\phi_1, \qquad \ldots\ldots(15)$$

we find from (12)

$$\begin{aligned}2\pi k\theta_0&=\int_{-\phi_0}^{\phi_0}\exp\{n(\phi-\phi_1)\}\{\mathrm{K}_0\,|n(\phi-\phi_1)|-\mathrm{K}_0(2n\alpha U)\}\{A\chi_0(\phi_1)+B\chi_1(\phi_1)\}\,d\phi_1\\&=A\left\{1-\mathrm{K}_0(2n\alpha U)\exp(n\phi)\int_{-\phi_0}^{\phi_0}\exp(-n\phi_1)\chi_0(\phi_1)\,d\phi_1\right\}\\&\quad+B\exp(n\phi)\left\{1-\mathrm{K}_0(2n\alpha U)\int_{-\phi_0}^{\phi_0}\exp(-n\phi_1)\chi_1(\phi_1)\,d\phi_1\right\}.\end{aligned} \qquad \ldots\ldots(16)$$

For (16) to be satisfied for all values of ϕ_1, A and B are given by

$$A=2\pi k\theta_0 \qquad \ldots\ldots(17)$$

$$B=\frac{A\mathrm{K}_0(2n\alpha U)\int_{-\phi_0}^{\phi_0}\exp(-n\phi_1)\chi_0(\phi_1)\,d\phi_1}{1-\mathrm{K}_0(2n\alpha U)\int_{-\phi_0}^{\phi_0}\exp(-n\phi_1)\chi_1(\phi_1)\,d\phi_1}. \qquad \ldots\ldots(18)$$

The heat loss H in unit time from the real plate for unit depth of the two-dimensional system may now be written down in terms of (14) and (15), viz.

$$\begin{aligned}\frac{H}{2\pi k\theta_0}&=\frac{1}{2\pi k\theta_0}\int_{-\phi_0}^{\phi_0}\chi(\phi_1)\,d\phi_1\\&=\int_{-\phi_0}^{\phi_0}\chi_0(\phi_1)d\phi_1+\frac{\mathrm{K}_0(2n\alpha U)\int_{-\phi_0}^{\phi_0}\exp(-n\phi_1)\chi_0(\phi_1)\,d\phi_1\int_{-\phi_0}^{\phi_0}\chi_1(\phi_1)\,d\phi_1}{1-\mathrm{K}_0(2n\alpha U)\int_{-\phi_0}^{\phi_0}\exp(-n\phi_1)\chi_1(\phi_1)\,d\phi_1}.\end{aligned} \qquad \ldots\ldots(19)$$

The problem is seen to reduce essentially to the solution of (14) and (15). That of the former equation, which refers to the distribution of heat sources appropriate to an isolated plate in a stream, has already been discussed (Piercy and Winny 1933). This paper gives for this part of the heat loss the alternative solutions

$$\int_{-\phi_0}^{\phi_0} \chi_0(\phi_1)\,d\phi_1 = -\frac{(n\phi_0)^2(\log\tfrac{1}{4}n\phi_0+\gamma+\tfrac{1}{2})+1}{\log\tfrac{1}{4}n\phi_0+\gamma},\quad n\phi_0 \text{ small} \quad \ldots\ldots(20)$$

$$\text{or } = 4(n\phi_0/\pi^3)^{1/2},\quad n\phi_0 \text{ large.} \quad \ldots\ldots(21)$$

γ is Euler's constant $=0{\cdot}5772\ldots$ The two results were shown to approach one another through an intermediate range of values of $n\phi_0$, of order unity, so that little numerical uncertainty exists at any value of this quantity. The distribution of heat sources appropriate to (20) is, from this earlier investigation,

$$\chi_0(\phi_1) = \frac{D+En\phi_1+F(n\phi_1)^2}{n({\phi_0}^2-{\phi_1}^2)^{1/2}}, \quad \ldots\ldots(22)$$

where the constants are defined by

$$\frac{n}{\pi D} = (\log\tfrac{1}{4}n\phi_0+\gamma)[\tfrac{1}{4}(n\phi_0)^2(\log\tfrac{1}{4}n\phi_0+\gamma-\tfrac{1}{2})-1],$$

$$E = D(\log\tfrac{1}{4}n\phi_0+\gamma+1),\qquad F=\tfrac{1}{2}D(\log\tfrac{1}{4}n\phi_0+\gamma+\tfrac{3}{2}), \quad \ldots\ldots(23)$$

whilst the distribution appropriate to (21) is

$$\chi_0(\phi_1) = \frac{1}{2}\left(\frac{2}{\pi}\right)^{3/2}\left(\frac{n}{\phi_0+\phi_1}\right)^{1/2}. \quad \ldots\ldots(24)$$

Before considering (15) we make use of the foregoing to evaluate a further integral which occurs in (19). From (22) and (23), for $n\phi_0$ small

$$\begin{aligned}\int_{-\phi_0}^{\phi_0}\exp(-n\phi_1)\chi_0(\phi_1)\,d\phi_1 &= \int_{-\phi_0}^{\phi_0}[1-n\phi_1+\tfrac{1}{2}(n\phi_1)^2]\frac{D+En\phi_1+F(n\phi_1)^2}{n({\phi_0}^2-{\phi_1}^2)^{1/2}}\,d\phi_1\\ &= \int_{-\phi_0}^{\phi_0}\frac{D+(E-D)n\phi_1+(\tfrac{1}{2}D-E+F)(n\phi_1)^2}{n({\phi_0}^2-{\phi_1}^2)^{1/2}}\,d\phi_1\\ &= \frac{\pi}{n}\{D+\tfrac{1}{2}(\tfrac{1}{2}D-E+F)(n\phi_0)^2\}\\ &= -[\log\tfrac{1}{4}n\phi_0+\gamma]^{-1}. \end{aligned} \quad \ldots\ldots(25)$$

From (24) for $n\phi_0$ large, writing m for $\{n(\phi_0+\phi_1)\}^{1/2}$,

$$\int_{-\phi_0}^{\phi_0}\exp(-n\phi_1)\chi_0(\phi_1)\,d\phi_1 = \left(\frac{2}{\pi}\right)^{3/2}\exp(n\phi_0)\int_0^{(2n\phi_0)^{1/2}}\exp(-m^2)\,dm. \quad \ldots\ldots(26)$$

Values follow from tables of the probability integral; the limiting value as $n\phi_0$ increases is $\frac{1}{2}\sqrt{\pi}$.

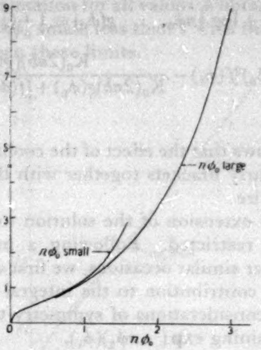


Figure 2. The integral $\int_{-\phi_0}^{\phi_0} \exp(-n\phi_0)\chi_0(\phi_0)\,d\phi_0$

The two approximations (25) and (26) are plotted in figure 2, where they will be seen to differ so little through a range of values of $n\phi_0$ of order 1 as to leave little numerical uncertainty at any value.

We proceed to examine (15) for $n\phi_0$ small. On expanding the Bessel function as far as the squares of small quantities the equation becomes, with r written for $\frac{1}{2}n(\phi-\phi_1)$,

$$1 = \int_{-\phi_0}^{\phi_0} \{-(\gamma+\log r)(1+r^2)+r^2\}\chi_1(\phi_1)\exp(-n\phi_1)\,d\phi_1. \quad \ldots\ldots(27)$$

Assume for the solution the form

$$\chi_1(\phi_1)\exp(-n\phi_1) = \frac{G+H\phi_1^2}{(\phi_0^2-\phi_1^2)^{1/2}} \quad \ldots\ldots(28)$$

and substitute in (27); the integrals can be evaluated exactly, and for the equation to be satisfied for all values of ϕ_1 to the second order of small quantities, we find

$$\begin{aligned} 1/G &= -\pi(\gamma+\log\tfrac{1}{4}n\phi_0)[1+(\tfrac{1}{2}n\phi_0)^2(\tfrac{3}{2}+\gamma+\log\tfrac{1}{4}n\phi_0)],\\ H &= \tfrac{1}{2}n^2G(\tfrac{1}{2}+\gamma+\log\tfrac{1}{4}n\phi_0). \end{aligned} \quad \ldots\ldots(29)$$

Still retaining the restriction of $n\phi_0$ small, we evaluate the two integrals in which the function (28) occurs. Firstly, we have

$$\begin{aligned} \int_{-\phi_0}^{\phi_0} \exp(-n\phi_1)\chi_1(\phi_1)\,d\phi_1 &= \int_{-\phi_0}^{\phi_0} \frac{G+H\phi_1^2}{(\phi_0^2-\phi_1^2)^{1/2}}\,d\phi_1\\ &= \pi(G+\tfrac{1}{2}H\phi_0^2). \end{aligned} \quad \ldots\ldots(30)$$

Secondly,

$$\begin{aligned} \int_{-\phi_0}^{\phi_0} \chi_1(\phi_1)\,d\phi_1 &= \int_{-\phi_0}^{\phi_0} \exp(n\phi_1)\frac{G+H\phi_1^2}{(\phi_0^2-\phi_1^2)^{1/2}}\,d\phi_1\\ &= -[\log\tfrac{1}{4}n\phi_0+\gamma]^{-1} \end{aligned} \quad \ldots\ldots(31)$$

to the second order the same result as (25).

This completes the working for $n\phi_0$ small, and under this condition the solution of (19) is as follows. Writing

$$f(\phi_0) \equiv \tfrac{1}{2} + \gamma + \log \tfrac{1}{4} n\phi_0, \qquad g(\phi_0) \equiv 1 + (\tfrac{1}{2} n\phi_0)^2 f(\phi_0), \qquad \ldots\ldots(32)$$

$$\frac{H}{2\pi k\theta_0} = \frac{1}{\frac{1}{2} - f(\phi_0)} \left\{ 1 + (n\phi_0)^2 f(\phi_0) - \frac{\mathrm{K}_0(2nb)[g(\phi_0) + (\frac{1}{2}n\phi_0)^2]}{\mathrm{K}_0(2nb) g(\phi_0) + [f(\phi_0) - \frac{1}{2}][g(\phi_0) + (\frac{1}{2}n\phi_0)^2]} \right\}. \qquad \ldots\ldots(33)$$

Comparison with (20) shows that the effect of the cooling boundary is represented by the last term in the curly brackets together with the variation of ϕ_0 from $2a$, its value for an isolated wire.

We now consider the extension of the solution to values of $n\phi_0$ beyond the range to which (33) is restricted. Following a method which has proved successful on several rather similar occasions, we first assume $n\phi_0$ to be large and, further, that all effective contribution to the integral of (15) then occurs when $\phi = \phi_1$. It is clear from considerations of symmetry that (15) is to be regarded as an equation for determining $\exp(-n\phi_1)(\phi_1)$.

Appropriately to the assumption now made, the equation becomes

$$1 = \exp(-n\phi)\chi_1(\phi) \int_{-\phi_0}^{\phi_0} \mathrm{K}_0(n\,|\phi - \phi_1|)\, d\phi, \qquad \ldots\ldots(34)$$

giving

$$\int_{-\phi_0}^{\phi_0} \exp(-n\phi)\chi_1(\phi)\, d\phi = \int_{-\phi_0}^{\phi_0} d\phi \left[\int_{-\phi_0}^{\phi_0} \mathrm{K}_0(n\,|\phi - \phi_1|) d\phi \right]^{-1}. \qquad \ldots\ldots(35)$$

By (10) the right-hand side is recast into the form

$$2\int_0^{\phi_0} n d\phi \left[\pi - \int_0^{\infty} \frac{\exp\{-n(\phi + \phi_0)\cosh t\} + \exp\{-n(\phi_0 - \phi)\cosh t\}}{\cosh t}\, dt \right]^{-1} \qquad \ldots\ldots(36)$$

in which it is observed that since $n\phi_0$ is large the first exponential is always negligible compared with the second. Also since

$$\int_{\xi}^{\infty} \mathrm{K}_0(\xi)\, d\xi = \int_0^{\infty} \frac{\exp(-\xi\cosh t)}{\cosh t}\, dt, \qquad \ldots\ldots(37)$$

(36) reduces to

$$2\int_0^{\phi_0} n d\phi \left[\pi - \int_{n(\phi_0 - \phi)}^{\infty} \mathrm{K}_0(\xi)\, d\xi \right]^{-1}. \qquad \ldots\ldots(38)$$

Hence writing

$$p \equiv n(\phi_0 - \phi), \quad I \equiv \int_p^{\infty} \mathrm{K}_0(\xi)\, d\xi \qquad \ldots\ldots(39)$$

and again identifying ϕ with ϕ_1,

$$I_1 \equiv \int_{-\phi_0}^{\phi_0} \exp(-n\phi_1)\chi_1(\phi_1)\, d\phi_1 = \frac{2}{\pi}\left\{ n\phi_0 + \int_0^{n\phi_0} \frac{I}{\pi - I}\, dp \right\}. \qquad \ldots\ldots(40)$$

The second term on the right-hand side requires to be evaluated graphically and it is found to approach the value $0{\cdot}271(\pi/2)$ as $n\phi_0$ increases.

The two solutions of the integral under discussion, viz. those given by (40) and (30) for $n\phi_0$ large and small, respectively, are compared in figure 3. It is inferred that a close approximation for all values is obtained by assuming $n\phi_0$ large to be greater than 1·5 and $n\phi_0$ small, less than 1 with the transitional curve shown dotted in the figure between these limits.

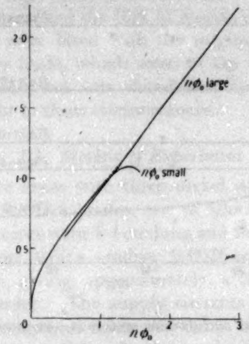


Figure 3. The integral $\int_{-\phi_0}^{\phi_0} \exp(-n\phi_1)\, \chi(\phi_1) d\phi_1$

The integral remaining before (19) can be evaluated for $n\phi_0$ large is

$$I_2 \equiv \int_{-\phi_0}^{\phi_0} \chi_1(\phi_1)\, d\phi_1. \qquad \ldots\ldots(41)$$

Multiplying both sides of (15) by $e^{n\phi}$ and retaining the first term only of the asymptotic expansion of the Bessel function, we have

$$e^{n\phi} = \int_{-\phi}^{\phi} \left[\frac{\pi}{2n(\phi-\phi_1)}\right]^{1/2} \chi_1(\phi_1)\, d\phi_1. \qquad \ldots\ldots(42)$$

Abel's solution then gives

$$\chi_1(\phi_1) = \left(\frac{2n}{\pi^3}\right)^{\frac{1}{2}} \frac{d}{d\phi_1} \int_{-\phi_0}^{\phi_1} \frac{e^{nt}}{(\phi_1 - t)^{1/2}}\, dt \qquad \ldots\ldots(43)$$

and for $n\phi_0$ large this reduces, leading to

$$I_2 \equiv \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \exp(n\phi_0) \int_0^{(2n\phi_0)^{1/2}} \exp(-t^2)\, dt, \qquad \ldots\ldots(44)$$

whose value for any value of $n\phi_0$ is immediately known.

Equation (27) now permits the heat loss from the plate to be evaluated with the help of figures 1, 2 and 3; an expression of this result in symbols, restricted so as to avoid the intervals of $n\phi_0$ where a transitional curve provides the best approximation possible, would be cumbersome, and so, too, would be its use.

Figure 4 exhibits the final solution of the problem, according to the present theory, giving, in the form of examples, the modification of the heat loss from the circular cylinder as it approaches the cooling plate. It will be seen that as speed increases, the cylinder may approach the plate more closely before excessive heat loss is experienced.

The assumption leading to equations (14) and (15) becomes untenable as a/b approaches unity and as Uan increases. This restriction leaves the theory useful over a wide range, including that of experimental explorations of velocity.

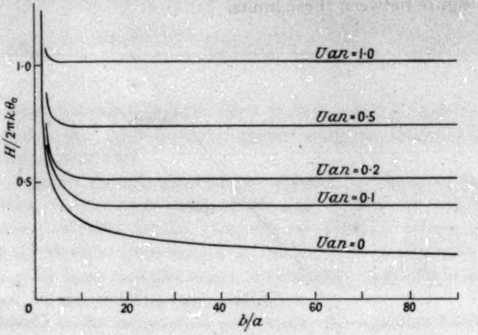


Figure 4. Theoretical non-dimensional heat loss at various values of Uan.

§ 2. EXPERIMENTAL

2.1. *Design of Apparatus*

To avoid turbulence the hot wires were required to be thin and to move at low speeds, while, according to § 1, their cooling would be expected to become excessive only when they approached a conducting body within distances comparable with their diameters. These conditions imposed a stringency on the design of the apparatus and the method of observation, in order that the small thickness of fluid separating wire from cooling surface might be set and gauged reliably.

A vertical shaft or axle, mounted between centres or in conical bearings in a heavy framework, was rotated through the reduction gearing by a 5 kilowatt electric motor driven from batteries to secure uniform speed. The axle carried two sliding frames, to each of which was attached an insulated blade forming an arm of the fork for the hot wire. These frames could be traversed horizontally a total distance of 3 cm by means of micrometer screws acting through the intermission of levers pivoted in knife edges and constraining strong springs, the double lever and micrometer system, operating each frame, providing coarse and fine adjustments. Each horizontal frame and its traversing gear formed a separate unit carried in a machined casting secured to the axle. The axle, the sliding frames and the hot wire formed a complete loop, so that the cooling body needed to be floated between the blades in ball bearings on the axle. This body consisted of a cylindrical drum 25 cm in diameter of polished brass. Tappets, tripped by a cam, kept the drum still, while permitting the hot wire fork to swing through. The stationary cooling surface thus provided was 5·4 cm long, but the surface was continued by stationary cylindrical casings of the same diameter and material fitted above and below so as to leave annular slots for the passage of the fork-blades. To minimize swirl the sliding frames and all other gear were enclosed in interior drums secured to the axle and rotating with it. Only the fork-blades projected through them. The outer surfaces of these moving drums were fitted concentrically, with small clearance, within internal cylindrical surfaces machined in the stationary drums. Thus little more than the fork-blades was left to create swirl, and these were shaped to a 'faired' section. The traversing screws were operated through doors in the casings.

Heavy copper leads connected the fork to mercury slip-rings at either end of the axle. Double rings were fitted with the original intention of connecting one pair to potentiometer leads, which were to tap the wire at some distance from its ends. The refinement was abandoned because of the difficulty of keeping a fine wire straight in these circumstances.

2.2. *Method of Experiment*

The experiments were made with three nickel wires of different diameters soldered tautly between the fork-blades, one of which was pivoted and sprung to provide tension. All wires were 5·4 cm long and they were heated as follows: radius 0·0013 cm, current 0·10 A, radius 0·0026 cm, current 0·40 A, radius 0·0038 cm, current 0·90 A, giving, approximately, a temperature of 40°C in still air remote from the cylinder. The supply currents were maintained constant during the experiments, resistance being allowed to change.

The procedure for a series of observations was as follows. The 'cold resistance' was determined by passing through a very small current, and the wire brought close to the cooling surface and set parallel to it by observing the wire and its optical image in the surface obliquely through a microscope. The value of the small clearance was determined by means of a calibrated eyepiece scale. The whirling-arm was then rotated at various speeds, which were noted by stop-watch, and the various resistances of the wire were determined by means of a simple high resistance bridge. The distance between wire and surface was then changed by a definite adjustment of the calibrated micrometer heads, while the microscope was again employed to check the new value of the clearance. A second set of resistances was then obtained, and the process repeated until the wire was too far removed from the cylindrical surface to be cooled by it even at low speed.

The linear velocities of the wires ranged from 5 to 60 cm sec^{-1}. The maximum speed gave for the wire of greatest diameter a Reynolds number less than one-tenth of the smallest value at which eddying has been observed behind long wires of circular section.

2.3. *Reduction of Results*

The hot resistance R_0 of a wire of length l and its cold resistance R_c are related to its temperature elevation θ_0 by

$$R_0 = R_c(I + \alpha\theta_0). \qquad \ldots\ldots(45)$$

The temperature coefficient of resistance α was specially determined by carrying out measurements on a coil of wire enclosed in a tube and immersed successively in ice, steam and boiling oil. Its mean value over the range of temperature provided was $4{\cdot}55 \times 10^{-3}$ deg^{-1}. H, the heat lost per second by unit length of the wire, was obtained from the heat supply in the steady state, so that

$$H = I^2R_0/Jl, \qquad \ldots\ldots(46)$$

I being the supply current in amperes and *J* the mechanical equivalent of heat. From (45) and (46), denoting by *k* the conductivity of the fluid,

$$\frac{H}{2\pi k\theta_0} = \frac{R_0 R_c}{R_0 - R_c}\,\frac{I^2\alpha}{2\pi J l k}. \qquad \text{......(47)}$$

Of these quantities, $R_0 - R_c$ was obtained directly in terms of the shift of the balance point along the bridge wire, while R_c, although also obtainable from the readings, was in preference determined more accurately by measuring the resistance of a number of specimens of the wire stretched on a holder of such a form that length could be varied. The mean value of the specific resistance of the nickel was found in this way to be for each diameter $8{\cdot}3 \times 10^{-6}$ ohm cm. Values of *k* were adopted from *International Critical Tables* to correspond with the mean temperatures of the wire and the cooling surface, the former being estimated from the observed resistance of the wire. Comparison with theory involves the quantity $n = c/2k$, and appropriate values of *c*, the heat capacity of the fluid, were taken from the tables with the same convention as to temperature.

For each wire and each experimental value of *b*/*a*, the clearance between wire and cooling surface expressed in terms of the radius of the wire, the observations of $H/2\pi k\theta_0$ were plotted against *Ua*. From the three families of curves so obtained, values of $H/2\pi k\theta_0$ were read off at chosen intervals of *Uan*. These are exhibited as points in figure 5, from which it will be observed that the wires gave results in

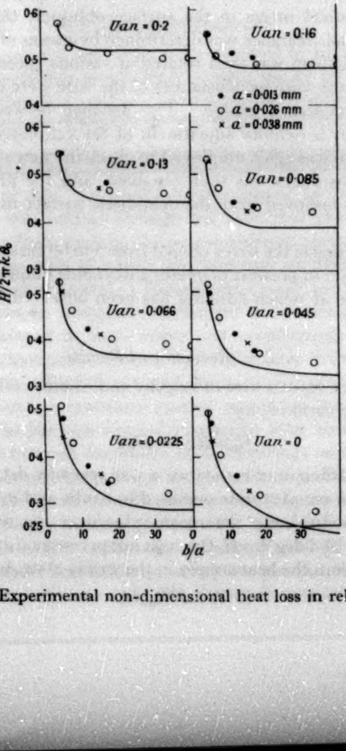


Figure 5. Experimental non-dimensional heat loss in relation to *b*/*a*.

agreement with one another. The curves shown in the figure are theoretical, but before making a comparison it is convenient to discuss briefly some sources of error and causes of divergence in the experiments which have not yet been considered.

3.4. *Sources of Error*

The theory does not take account of radiation losses. In the case of the thickest wire these were estimated to be 3×10^{-3} cal cm^{-1} sec^{-1} under the conditions of the experiments, and the correction was consequently neglected.

Natural convection, which may be expected to increase the heat loss at very small velocities, is also ignored in the theory. An estimate of its importance was sought by mounting a specimen of the wire vertically in a wide vertical tube connected to an aspirator, which could produce a downward flow of air in the tube counter to the upward current caused by free convection. It was found that for the wire to attain a maximum temperature a downward current of some 2 cm sec^{-1} was required. Thus observations of heat loss with the wire at rest correspond to the existence of an appreciable velocity parallel to its axis, but the effect on the results at normal velocities of the order of 5 cm sec^{-1}, or greater, will be small.

Departure in the experiments from two-dimensional conditions also becomes important at low speeds. If the wires be assumed to conform in shape to long ellipsoids, the steady conduction of heat under static conditions may readily be calculated. This was found to be appreciable in the experimental circumstances, but the result illustrates a well-known difficulty arising from the two-dimensional assumption in the steady state, and the theory must be supposed, on this score alone, to break down as the velocity approaches zero.

Error arises in two ways from the leads and soldered electrical junctions. Their resistance does not affect $R_0 - R_c$, but it prevents direct measurement of the absolute value of either quantity separately. As already mentioned, the difficulty was met by inferring R_c from special experiments of a more general kind. The other error is due to conduction of heat along the metal blades of the fork, and a direct estimate under present conditions was thought to be advisable. Calibration curves for the hot wires were obtained in an aspirator tube with the fork formed (*a*) of thick copper blades immersed in ice, (*b*) of thin needles in air. Differences amounted to less than 3%.

Preparations were made to correct the timed velocity of the wire for a velocity of swirl set up by the rotating parts of the whirling-arm apparatus, a stationary hot wire being mounted near the path of the moving wire. The steps taken in the design of the apparatus specially to minimize this drift of air proved effective, however, for no appreciable swirl could be detected.

The most serious cause of discrepancies in early observations was found to lie in estimating the clearance between the wires and the boundary. Fair accuracy was clearly obtained in any one measurement by use of the method which has been described, and the cause of error was finally traced to a slight eccentricity of the axle within the drum. The difficulty was overcome by choosing a suitable particular range of angular position on the wire both for measurements of clearance and for observations of heat loss.

3.5. *Comparison with Theory*

In figure 5 the curves are calculated from § 1 and the points represent the present experiments: fairly close agreement is seen to exist.

At all but the greatest values of *Uan* there will be noticed a systematic divergence between experiment and theory, the latter tending to under-estimate the heat loss. The difference was investigated on the experimental side in the light of the sources of error detailed in the preceding section, and was found to be genuine. It might have been predicted. The theory of convection from isolated wires in an inviscid stream advanced by two of the present authors was shown (Piercy and Winny 1933) to indicate a smaller heat loss at values of *Uan* less than 0·2 than was indicated by a rearranged record of King's experiments on isolated wires moving through air. The difference there found was of the same order as that under present discussion.

The foregoing comparison suggests a method by which the theory of § 1 may be employed with accuracy. The heat loss from an isolated hot wire is readily determined experimentally, while the excess heat loss due to the wire approaching a conducting boundary may be calculated from § 1 in terms of the isolated heat loss and may be applied in that form to the experimental calibration. Figure 6 illustrates the results finally obtained. The broken lines represent the present experiments and the full lines are calculated from the theory, except that they are constrained in the manner described to agree with experiment at $b/a = \infty$. Small differences remain but, having regard to difficulties of accurate observation, the theoretical prediction is of equal reliability.

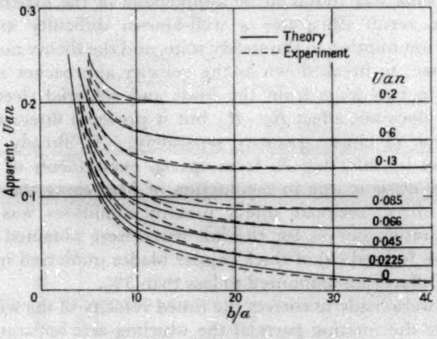


Figure 6. Apparent *Uan* in relation to *b*/*a*.

REFERENCES

KING, L. V., 1914, *Phil. Trans. Roy. Soc.* A, **214.**
PIERCY, N. A. V., and WINNY, H. F., 1933, *Phil. Mag.*, **16**, **390.**

The Electrical Conductivity of Hexagonal Selenium

By R. A. HYMAN†
Physics Department, University of Reading

Communicated by R. W. Ditchburn; *MS. received* 16*th December* 1955 *and in revised form* 30*th January* 1956

Abstract. A model of selenium is discussed in which it is proposed that a crystal is formed on continuous chains of atoms which zigzag so as to fill a volume which is large compared with the lattice spacing. It is assumed that the charge carriers (holes) move freely along these chains but that transitions between chains are difficult. This predicts the right dependence of conductivity on frequency and a mobility rising with temperature but the magnitude of the resistivity is not explained since this requires a detailed discussion of the transition process. Some related properties are discussed.

§ 1. INTRODUCTION

During recent years it has been shown that the electrical properties of selenium are unusual. The mobility of the positive holes, which are the majority carriers, is very small, of the order of $0{\cdot}1\,\mathrm{cm}^2\,\mathrm{v}^{-1}\,\mathrm{sec}^{-1}$ at room temperature, and increases with temperature. The conductivity is dependent on both the frequency (Henkels 1951 b, Rebstock and Seiler 1954) and on the magnitude (de Boer 1947, Plessner 1951 a, Henkels 1951 b) of the applied field. Attempts to explain these features in terms of barrier layers have not been satisfactory since the unusual features occur both in microcrystalline selenium and in single crystals where there seems to be no physical basis for the barrier layers. Further, the mobility of carriers in the isomorphous element tellurium, and in selenium–tellurium alloys containing up to 13% selenium by weight, is high, about $1000\,\mathrm{cm}^2\,\mathrm{v}^{-1}\,\mathrm{sec}^{-1}$, and decreases with rising temperature in the normal metallic fashion (Nussbaum 1954). It thus seems desirable to attempt to derive the properties of selenium from its peculiar chain structure and to determine the cause of the qualitative difference between the electrical properties of selenium and tellurium.

The transition of carriers between separate molecules of selenium may be compared to the transitions between localized states (de Boer and Verwey 1937) in, for example, NiO, Fe_2O_3 and Fe_3O_4, in which the electrons must be considered as localized at the individual ions: whereas in these oxides transitions occur between neighbouring ions, in selenium the rate-determining processes are transitions between adjacent molecules. In selenium and Fe_3O_4 (Morin and Geballe 1955) the activation energy associated with the mobility μ is 0·1 ev; in α-Fe_2O_3 and NiO (Morin 1954) the activation energy is approximately 0·8 ev, falling to 0·1 ev at a sufficiently high impurity concentration or temperature.

---

† Now at Standard Telecommunication Laboratories Ltd., Enfield, Middx.

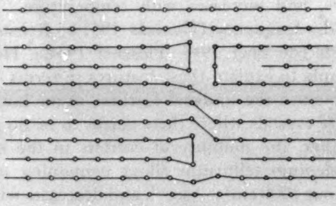
The transition probability is proportional to the diffusion constant for the carriers, $D=\mu kT/q$, where q is the charge of the carriers.

Potential barriers of thickness of the order of the interatomic spacing and a fraction of an electron volt in height should be transparent to the electrons, but it may tentatively be suggested that the need for transitions between localized states to be accompanied by a reaction on the lattice is the origin of the exponent in the expression for the transition probability, and of the low mobility. Consider, for example, the case of the transition of an electron from one state to a neighbouring vacant state of equal energy. The transition of an electron by itself cannot be a complete description of the process since, if it were, the centre of mass of the system would have moved. Therefore the transition must be accompanied by a reaction on the lattice which will require the initial absorption of energy, whatever the exact form of the interaction.

In what follows it is simply assumed that the rate determining process in the transition of carriers between selenium chains is associated with an activation energy, without further consideration of its origin.

§ 2. STRUCTURE

The chains in hexagonal selenium have the form of spirals parallel to the *c*-axis and held together by residual forces. These spirals can fold back along a neighbouring position in the lattice with little distortion of the covalent bond lengths and angles and, since the interchain forces are weak, the distortion of the chains caused by the folding requires little energy. With imperfections of this type the chains would zigzag through the lattice (see figure). Let us call a length of this chain parallel to the *c*-axis a 'link', and each other section a 'cross link', so that each length of chain, or each molecule, is formed of a number of links each formed of atoms bound together by single covalent bonds. If the chain is terminated by a selenium atom—and not by a halogen impurity atom—the end has an unsaturated valency which, it may be assumed, forms an acceptor.



Schematic drawing of extended selenium chain.

To simplify discussion let each chain extend through a rectangular solid, and let these rectangular solids be of uniform shape and size. Let one side of the solid of length $d_{\parallel}$, be parallel to the *c*-axis and the other two sides have equal length, written $d_{\perp}$. As a result of the zigzag structure, each of these dimensions is large compared with the lattice spacing.

§ 3. Electrical Model

We now derive the frequency dependence of the conductivity parallel to one of the sides of the rectangular solid with side of length d. Let the voltage drop across d be V'. As a first approximation take the current between molecules to be zero, so that the conduction current is everywhere balanced by an equal diffusion current. This approximation will be valid if the resistivity within each molecule is sufficiently small.

Let n be the density of carriers with equilibrium value n_0, and the voltage $V=0$ where $n=n_0$. Since there is equilibrium, $n=n_0\exp(-qV/\mathbf{k}T)$. From Poisson's equation,

$$-d^2V/dx^2=(4\pi q/\epsilon)(n-n_0),$$

or

$$d^2V/dx^2=(4\pi qn_0/\epsilon)\{1-\exp(-qV/\mathbf{k}T)\} \qquad \ldots\ldots(1)$$

where q is the charge of a hole and ϵ is the dielectric constant. Put $4\pi q^2n_0/\epsilon\mathbf{k}T=\alpha^2$; then, for $|V|\ll\mathbf{k}T/q$, $d^2V/dx^2=\alpha^2$. Take $x=0$ at $V=0$. The charge in each region is conserved, $n-n_0\propto V$, and V is a symmetrical function of x so that $V=V'/2$ at $x=d/2$.

Therefore

$$V=\frac{V'\sinh\alpha x}{2\sinh\alpha d/2}.$$

With a typical value of $n_0=10^{14}\,\mathrm{cm}^{-3}$, $\alpha=3{\cdot}4\times10^4\,\mathrm{cm}^{-1}$. Also, the total charge flowing to one side of the region, per unit cross-sectional area,

$$Q=\int_0^{d/2}(n-n_0)\,dx=n_0\frac{q^2V'}{2\alpha\mathbf{k}T}\,\frac{\cosh(\alpha d/2)-1}{\sinh(\alpha d/2)}.$$

A system, of unit length, of these regions may be regarded as $1/d$ condensers in series, each having a charge Q; and the total voltage across the system is $(1/d)V'$, so that the capacity per unit length

$$C=\frac{dQ}{V}=n_0\frac{dq^2}{2\alpha\mathbf{k}T}\,\frac{\cosh(\alpha d/2)-1}{\sinh(\alpha d/2)}. \qquad \ldots\ldots(2)$$

To calculate the current it is necessary to take a finite value for the transition probability. Let τ be the average time a hole spends in a molecule before moving a neighbouring link of another molecule. The net current per unit area $I=(n_1-n_2)bqx'/\tau$, where subscripts 1 and 2 refer to opposite sides of a boundary or opposite ends of a region, b is the thickness of a layer of atoms, and x' is a numerical factor of order unity. If, for example, the regions are divided by planes parallel to the c-axis, $x'=\frac{1}{3}$.

Integrating equation (1),

$$\frac{1}{2}\left(\frac{dV}{dx}\right)^2=\frac{4\pi qn_0}{\epsilon}\left[V-\frac{\mathbf{k}T}{q}\exp\left(-\frac{qV}{\mathbf{k}T}\right)\right]+K.$$

But dV/dx is continuous, so that $(dV/dx)_1=(dV/dx)_2$. Therefore

$$\begin{aligned}V'=V_2-V_1&=(\mathbf{k}T/q)\{\exp(-qV_1/\mathbf{k}T)-\exp(-qV_2/\mathbf{k}T)\}\\&=(\mathbf{k}T/q)(n_1-n_2)/n_0,\\I&=(n_0bx'/3\tau)(q^2V'/\mathbf{k}T),\end{aligned}$$

and the resistivity $\rho=(3\tau/n_0bdx')(\mathbf{k}T/q^2)$. ......(3)

The dependence of the resistivity on temperature appears as the exponential dependence of τ on temperature. From equations (2) and (3), the time constant of the crystal,

$$\rho C = \frac{3\tau}{2\alpha b x'} \frac{\cosh(\alpha d/2) - 1}{\sinh(\alpha d/2)} . \qquad \ldots\ldots(4)$$

Now it is found experimentally that all the acceptors are fully ionized, so that in the pure material, since each molecule has two ends, $d_{\perp}{}^2 d_{\parallel} = 2/n_0$. Assume that $\rho_{\perp} d_{\perp} \sim \rho_{\parallel} d_{\parallel}$ and take experimental values for the resistivity, $\rho_{\perp} \sim 10^6$ ohm cm and $\rho_{\parallel} \sim 10^5$ ohm cm (Plessner 1951 a), then $d_{\perp} \sim 10^{-5}$ cm, and $d_{\parallel} \sim 10^{-4}$ cm. From equations (3) and (4), we then find $\rho_{\parallel} C_{\parallel} \sim 10^{-7}$ sec and $\rho_{\perp} C_{\perp} \sim 10^{-8}$ sec.

The only published measurements of the frequency dependence of the conductivity have been made on microcrystalline selenium with a carrier concentration, $n_0 \sim 10^{15}$ cm^{-3}, in which it was found that the conductivity was independent of frequency at frequencies below the order of 10^7–10^8 c/s; at higher frequencies the conductivity was found to increase rapidly with rising frequency. This behaviour can be represented approximately by a resistance–capacity network with a time constant $\rho C \sim 10^{-7}$–10^{-8} sec. Therefore the theory given above accounts for the rapid observed increase of conductivity as a function of frequency, and gives the order of magnitude of the frequency at which this increase is found.

§ 4. MOBILITIES WITHIN A MOLECULE

In order to consider the assumption made above that the bulk of the resistance occurs between separate molecules we next estimate the mobility of the carriers within a single molecule. We are here concerned with the lengths of individual links of a chain.

Let $l_{\parallel}$ be the average length of a 'link' parallel to the c-axis, $l_{\perp}$ the average length of a 'cross link', and v the average velocity of a hole.

If the extension of the chain through the lattice follows a 'random walk' and a hole moves freely along the chain, the mobility parallel to the c-axis $\mu_{\parallel}' = qD_{\parallel}'/\boldsymbol{k}T = qvl_{\parallel}/2\boldsymbol{k}T$. Similarly $\mu_{\perp}' = qD_{\perp}'/\boldsymbol{k}T = (vl_{\perp}{}^2/l_{\parallel})(q/4\boldsymbol{k}T)$, the factor $\frac{1}{2}$ being due to planar geometry.

Take $l_{\parallel} = 10^{-6}$ cm and $l_{\perp} = 4{\cdot}4 \times 10^{-8}$ cm, the interchain distance. Assuming the effective mass of a hole to be that of a free electron, $v = 6 \times 10^6$ cm sec^{-1} at room temperature so that $\mu_{\parallel}' = 100$, and $\mu_{\perp}' = 0{\cdot}1$ cm^2 v^{-1} sec^{-1}. These may be compared with Plessner's (1951 a) experimental values for the mobility in the material as a whole, $\mu_{\parallel} = 0{\cdot}1$, $\mu_{\perp} = 0{\cdot}01$ cm^2 v^{-1} sec^{-1}. It has been assumed above that each hole reaching the end of a link moves to the succeeding link in the chain. In fact some of the holes will be reflected, but it may safely be concluded that the approximation is valid parallel to the c-axis. Perpendicular to the c-axis the approximation may introduce some error at room temperature but improves as the temperature is reduced. In both cases, parallel and perpendicular to the c-axis, most of the resistance occurs at the boundaries between separate molecules.

§ 5. RELATED STRUCTURES

5.1. *Liquid Selenium*

Liquid selenium is a p-type conductor with the number of carriers n proportional to $\exp(-E/2\boldsymbol{k}T)$, where $E = 2{\cdot}3$ ev (Lizell 1952). This behaviour is not easy to understand in terms of activation across a forbidden band, but it follows from the assumption made above that the free chain ends form acceptor centres, as the Se–Se bond energy is of the order of 2·5 ev.

5.2. *Graphite*

In relation to the assumption that the charge carriers in selenium move freely only along the chains, it is interesting to note that in the planar structure graphite, the experimental results for the conductivity of single crystals (Dutta 1953) are consistent with the interpretation that the carriers move freely only along the planes, the conductivity perpendicular to the planes being due to misorientations in the planes in the mosaic structure found in most real crystals.

5.3. *Tellurium*

Tellurium and selenium–tellurium alloys containing up to 13% selenium by weight, which are isomorphous with hexagonal selenium, do not have the anomalous electrical properties of selenium. If this difference is a result of the holes in selenium being localized in the chains, then presumably at a sufficiently high pressure the conductivity of selenium must become of the same type as that of tellurium. Measurements of the conductivity of selenium seem only to have been made at pressures of up to 3000 kg cm^{-2} (Montén 1909). In order to create in selenium interchain forces of the order of the van der Waals forces in tellurium, pressures of the order of 30 000 kg cm^{-2} would be necessary.

§ 6. CONCLUSION

In order to determine the magnitude of the conductivity a calculation of transition probability for charge transfer between chains would be necessary. From the hypothesis that the carriers in selenium move freely only along the chains together with certain structural hypotheses, an explanation has been given of the frequency dependence of the conductivity of hexagonal selenium.

ACKNOWLEDGMENTS

The author wishes to thank Dr. R. J. Elliott, Dr. E. W. J. Mitchell and Dr. W. N. Reynolds for many useful discussions and Standard Telecommunication Laboratories for leave of absence and support.

REFERENCES

DE BOER, F., 1947, *Philips Res. Rep.*, **2**, 349.

DE BOER, J. H., and VERWEY, E. J. W., 1937, *The Conductivity of Electricity in Solids*, *Proc. Phys. Soc.* (Extra Part) **49**, 59, 72.

DUTTA, A. K., 1953, *Phys. Rev.*, **90**, 192.

HENKELS, H. W., 1951 a, *J. Appl. Phys*,. **22**, 916 ; 1951 b, *Ibid.*, 1265.

LIZELL, B., 1952, *J. Chem. Phys.*, **20**, 672.

MONTÉN, F., 1909, *Dissertation*, Uppsala.

MORIN, F. J., 1954, *Phys. Rev.* **93**, 1195, 1199.

MORIN, F. J., and GEBALLE, T. H., 1955, *Phys. Rev.*, **99**, 467.

NUSSBAUM, A., 1954, *Phys. Rev.*, **94**, 707.

PLESSNER, K. W., 1951 a, *Proc. Phys. Soc.* B, **64**, 671 ; 1951 b, *Ibid.*, **64**, 681.

REBSTOCK, H., and SEILER, K., 1954, Z. *Naturf.*, **9a**, 49.

The Discharge Mechanism in the High-Vacuum Cold-Cathode Pulsed X-ray Tube

By P. T. G. FLYNN†

Nuclear Particle Laboratory, Queen Mary College, University of London

MS. received 4th June 1955, *and in revised form 24th February* 1956

Abstract. The existing theory of the discharge mechanism in the cold-cathode pulsed x-ray tube based on ionization of residual gas and plasma formation is reviewed and is shown to fail for a high-vacuum tube. A new theory is presented which is based on the production at the cathode of a vapour jet having a velocity of approximately 10^6 cm sec^{-1}. This results in an expanding region of low voltage discharge, and leaves most of the applied voltage across a rapidly contracting vacuum region adjacent to the anode. It is shown that the tube current and voltage as functions of time and the x-ray pulse duration can be predicted from a knowledge of the voltage supply and tube parameters, and the results are verified experimentally.

§ 1. INTRODUCTION

THE simplest method of producing a short pulse of x-rays is to discharge a capacitor through a hot-cathode x-ray tube (Oosterkamp 1940, Zuckerman 1943), and high speed x-ray motion pictures can be taken in this way (Dickson, Zavales and Ehrke 1948, Slack, Ehrke, Zavales and Dickson 1949 a, b). For higher voltages either a normal impulse generator or a delay-line impulse generator (Clayton 1952) is used. Discharge times, however, are limited to a useful range above about 10 μsec owing to the difficulty of obtaining from a hot cathode the large current required.

By using a tube containing mercury vapour, either with a hot cathode (Steenbeck 1938) or with a cold cathode (Kingdon and Tanis 1938), shorter discharge times of the order of several microseconds have been obtained, but the shortest discharge times, down to several tenths of a microsecond, have been achieved with cold-cathode high-vacuum tubes using an arc at the cathode for discharge initiation (Slack and Ehrke 1941). These tubes have been considerably developed, both in the U.S.A. (Slack and Dickson 1947, Slack and Ehrke 1948, Clark 1947 a, b, 1949, Breidenbach 1949, Criscuolo and O'Connor 1953) and in Germany (Simon 1947, Schaafs 1949, Fünfer 1950, 1953, Thomer 1953, Schall 1953, Schaafs and Herrmann 1954).

At the highest voltages both hot- and cold-cathode high-vacuum tubes give fairly long pulses (Brasch and Lange 1930, 1936, 1937, Brasch 1947, Sievert 1950, Lindell, Sievert and Wahlberg 1950).

The advantage of the cold-cathode tube is that no cathode supply is required. However, it is well known that in comparison with the hot-cathode type such tubes are not stable and are difficult to operate; for this reason information on the mechanism of the discharge in the cold-cathode tube is useful and may assist design which at present is based on empirical principles.

---

† Now at the Atomic Weapons Research Establishment, Aldermaston, Berks.

Theories exist regarding the mechanism of the discharge which can account for the current and voltage as functions of time within a certain pressure range, but they appear to fail for very low pressures. It is the purpose of this paper to analyse the difficulties of the existing theories and to propose an alternative mechanism applying to the high-vacuum cold-cathode pulsed x-ray tube. The theoretical argument is verified experimentally.

The type of tube considered here is the cold-cathode pulsed x-ray tube (see figures 5 and 6) in which a field emission arc initiated at the cathode leads to a discharge which results in emission of x-rays from the anode.

§ 2. THE MECHANISM OF THE DISCHARGE

The only theoretical work on the discharge process is that due to Steenbeck (1938) and although Fünfer (1950) has extended the theory it remains essentially unchanged. Various other suggestions regarding the mechanism have been made in the literature without, however, any formulation into a usable theory.

The basic task of any theory of the discharge is to explain the way in which the current through the tube is enabled to grow, since it is well known that the peak discharge current exceeds by some orders of magnitude the pure electron space-charge-limited current. There are two ways in which the space-charge limitation can be overcome; the first, adopted by Steenbeck and Fünfer, is that positive ions neutralize the electron space charge at the cathode; the second, which is proposed in this paper, is that the distance over which electron space charge limits the current is reduced.

2.1. *Theory of Steenbeck and Fünfer*

Steenbeck (1938) suggested that electrons from the cathode ionize the residual gas in the tube, producing positive ions which are effectively stationary in the short time of discharge, and by reducing the electron space charge allow larger currents to flow, the effect being cumulative. Assuming that at any time approximately equal numbers of ions and electrons are present Steenbeck obtained for the tube current i the relation:

$$i = Qsp(\mathbf{e}/\mathbf{m})^{1/2}(V_0 - Q/C)^{1/2} \qquad \ldots\ldots(1)$$

where Q is the total charge that has passed at any time, s is the number of positive ions produced per unit path length in gas at unit pressure, p is the pressure, $\mathbf{e}$ and $\mathbf{m}$ are the charge and mass of the electron, and V_0 is the voltage to which the capacitance C is initially charged.

Fünfer (1950) modified Steenbeck's theory to take account of the variation of the ionization function s with electron energy and obtained for the tube current:

$$i = -pC(2\mathbf{e}V/\mathbf{m})^{1/2}\int_{V_0}^{V} s\,dV. \qquad \ldots\ldots(2)$$

The tube voltage can be found from the relation

$$i = -C\,dV/dt \qquad \ldots\ldots(3)$$

and in figure 1 is shown the result of a calculation made by Fünfer. Experiments to check the theory using the capacitance and voltage for which the curves of figure 1 were calculated gave a maximum current of 50 A and a discharge duration of 1·5 μsec; the theoretical values are 2 A and 30 μsec.

The clue to the reason for the failure of the theory may be found in figure 2 which compares Steenbeck's results for maximum current with the measurements of Kingdon and Tanis. It can be seen that at the lower pressures the maximum current is no longer proportional to pressure but is roughly constant with a magnitude greater than the theoretical value. Fünfer's measurements were made at a pressure of 3×10^{-5} mm Hg, significantly lower than the region in which Steenbeck obtained agreement with theory, and again the measured current is higher than the theoretical value. It is thus apparent that the theories of Steenbeck and Fünfer fail because the mechanism of gaseous ionization can no longer account for the production of sufficient positive ions. Fünfer suggests that the extra positive ions may be produced by photoionization. Little information is available on this possibility, but it seems unlikely that at the pressures used the cross section would be large enough to give the required increase in current. In any case, if this mechanism were effective it would operate also at higher pressures and thus maximum currents proportional to pressure would still be obtained; this is found not to be true. The theory is also not satisfactory in view of certain other experimental results which will be discussed later. An alternative mechanism is therefore required, and one will now be proposed that will apply to the high-vacuum tube in which gaseous ionization is negligible.

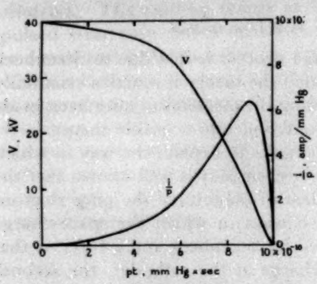


Figure 1. Fünfer's theoretical curves for the tube current and voltage for a 0·0005 μF capacitor charged to 40 kv.

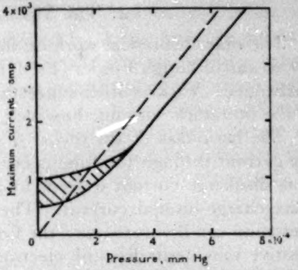


Figure 2. Maximum current in the discharge as a function of pressure. Full line: Experimental results of Kingdon and Tanis. Broken line: Steenbeck's theoretical curve.

2.2. *An Alternative Theory*

If the possibility of ionization of residual gas is rejected, other sources of positive ions must be considered in order to explain the large magnitude of the peak current in the discharge in relation to the pure electron space-charge-limited current. The following are possible sources of positive ions: (*a*) the electrodes: because of the field direction only the anode need be considered, and an immediately obvious means of production is electron bombardment; (*b*) the tube walls: again electron bombardment is an obvious means of production; (*c*) evolved gas: this could come from either electrode or from the tube walls and it then would be ionized in the same way as residual gas.

For the purpose of this analysis any process involving the tube walls will be rejected. Although such processes may be important in certain cases it is possible to make their effect small by ensuring that the tube walls are far away compared with the inter-electrode distance.

It should be noted at this point that no breakdown between the tube anode and cathode will occur in the absence of the triggering discharge.

2.2.1. *Production of positive ions at the anode and space-charge limitation.*

The process of production of positive ions by electron bombardment of the anode, however efficient, is not by itself sufficient to account for the large currents occurring in the discharge. This may be shown by considering a result due to Langmuir (1929), namely that even with infinite supplies of electrons at the cathode and positive ions at the anode the maximum total current is

$$i_{\max} = 1{\cdot}86(i_e + i_p) \qquad \ldots\ldots(4)$$

where i_e and i_p are the pure electron and the pure positive ion space-charge currents respectively. Now for given conditions the space-charge-limited current is inversely proportional to the mass of the charge carriers, assuming that they are singly charged, and therefore

$$i_p/i_e = (m_e/m_p)^{1/2} \qquad \ldots\ldots(5)$$

where m_e and m_p are the masses of electrons and positive ions respectively. The maximum current thus becomes

$$i_{\max} = 1{\cdot}86[1 + (m_e/m_p)^{1/2}]i_e. \qquad \ldots\ldots(6)$$

It can be seen that even with the lightest available ions the formation of positive ions at the anode cannot explain the observed current magnitudes. This conclusion implies that any ionization to be effective must take place at points between the electrodes and not at electrode surfaces.

2.2.2. *Ionization resulting from gas evolved at the electrodes.*

It is now necessary to consider whether the required ionization results from gas which is evolved at the electrodes and flows into the gap. This mechanism is at first sight a very attractive one, but consideration of the experimental evidence leads to difficulties in its acceptance.

Some results bearing on the case considered here have been obtained by Allwood.† In this work rectangular voltage pulses of 200 kv with a duration of 5 μsec were applied to a hot-cathode x-ray tube. A corresponding rectangular current pulse of 10 A was obtained and it is significant that in almost all cases the current remained at 10 A for the duration of the pulse, i.e. the anode could sustain bombardment of this magnitude over a fairly small area without supplying gas for ionization within the gap.

Dickson, Zavales and Ehrke (1948) found that a current of 60 A from a hot cathode was maintained with only a slight increase when a capacitor charged to 117 kv was discharged through the tube. In this case the discharge duration was 25 μsec.

---

† The author is indebted to the Director of Research, Nelson Research Laboratories, English Electric Co. Ltd., for permission to quote unpublished results.

Several authors attribute the current magnitude to vapour produced by heating of the anode, but it is difficult to accept this suggestion, since in order to produce sufficient heat for vaporization a large current must already be flowing.

It therefore seems justified to conclude that production of gas at the electrode does not account satisfactorily for the discharge.

2.2.3. *The triggering arc.*

Much light can be thrown on the mechanism of the discharge by consideration of the triggering arc. In the tube shown in figure 5 the sharp edges of the cathode and third electrode are separated by less than a tenth of a millimetre, and on application of the voltage breakdown occurred in a time too short to be measured with the present equipment. Hull and Burger (1928) and Snoddy (1931) found that transition to arc conditions for similar geometries occurred in times of less than 10^{-7} sec. This time is sufficiently short to allow the assumption that an arc is running at the cathode of the tube for the duration of the main discharge.

Experiments made by several investigators show that an arc between electrodes in a vacuum produces a high velocity vapour jet characteristic of the electrode material (Tanberg 1930, Kobel 1930, Easton, Lucas and Creedy 1934). There is thus a highly effective means available for producing vapour between the electrodes in the type of tube considered here. All the velocities measured by Easton, Lucas and Creedy are of the order of 10^6 cm sec^{-1}. This is a sufficiently high velocity for the vapour to travel well into the gap during the discharge, and there to be ionized. Alternatively, since the vapour is produced by the triggering arc, it is possible that it is already ionized, and that the jet actually consists of a mixture of electrons and positive ions. Either possibility allows the current and voltage for the discharge to be calculated.

2.2.4. *Calculation of current and voltage as functions of time.*

The mechanism of the discharge postulated here is as follows. On applying voltage to the three-electrode tube intense field emission between the cathode and the auxiliary electrode leads in a very short time to the initiation of the triggering arc, which causes a jet of vapour to be sent at high velocity from the cathode into the space between the anode and cathode. It is then assumed that the gap may be divided into two regions. The first extends from the anode to the front of the advancing vapour jet and is a vacuum region in which no ionization takes place. The second region extends from the cathode to the front of the advancing jet and will be assumed to be effectively conducting so that all the applied voltage appears between the advancing vapour front and the anode. The limitation on the current will be that set by electron space-charge in the vacuum region, and it will be assumed that this current can be drawn from the advancing front.

On the basis of these assumptions the current and voltage waveforms may be easily calculated. Consider electrodes of area A separated by distance d in vacuum. These are connected to a capacitance C charged to V_0, and at time $t=0$ the triggering arc occurs at the cathode and a conducting plasma begins to travel towards the anode with velocity v. The effective gap at any time will be $d-vt$ and, assuming electron space-charge limitation, the equation for the current i may be written

$$i=kAV^{3/2}(d-vt)^{-2} \qquad \text{......(7)}$$

where V is the tube voltage and k is a constant. Since $i=-C\,dV/dt$ the equation becomes

$$-C dV/dt = kAV^{3/2}(d-vt)^{-2}. \qquad \ldots\ldots(8)$$

Inserting the boundary condition $t=0$ when $V=V_0$ we find for the voltage

$$V=4C^2d^2V_0(d-vt)^2[2Cd^2+(kAV_0^{1/2}-2Cdv)t]^{-2} \qquad \ldots\ldots(9)$$

and for the current

$$i=8kAC^3d^3V_0^{3/2}(d-vt)[2Cd^2+(kAV_0^{1/2}-2Cdv)t]^{-3}. \qquad \ldots\ldots(10)$$

Equations (9) and (10) are plotted in figure 3. Representative curves are shown for the discharge of a 0·01 μF capacitor charged to 100 kv. The constant k has a value 2·34 A V$^{-3/2}$. Easton, Lucas and Creedy (1934) give values for v; in view of the dependence of r.m.s. velocity on the type of distribution assumed and the fact that no account has been taken of any distribution of velocities in deriving equations (9) and (10), a value for v of 10^6 cm sec^{-1} has been used in calculating the curves. It should be noted that production of positive ions at the anode can be allowed for by replacing A by ϕA where ϕ is a numerical factor having a maximum value $1{\cdot}86[1+(m_e/m_p)^{1/2}]$ (see equation (6)). The value of ϕ for any given production of ions may be found by reference to Langmuir (1929).

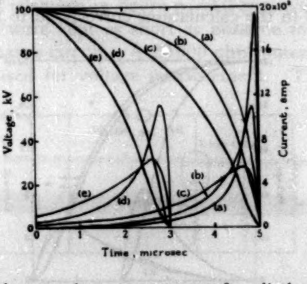


Figure 3. Theoretical voltage and current curves for discharge with no expansion. $C=0{\cdot}01\,\mu$F, $V_0=100$ kv. (*a*) $A=4$ cm², $d=5$ cm. (*b*) $A=8$ cm², $d=5$ cm. (*c*) $A=16$ cm², $d=5$ cm. (*d*) $A=8$ cm², $d=3$ cm. (*e*) $A=16$ cm², $d=3$ cm.

It can be seen from figure 3 that the peak current in the discharge is greatly affected by the value of A, but that alteration of d produces relatively little effect. This is because the period during the close approach of the conducting region to the anode is more important than earlier periods in the discharge; only then can sufficiently large currents flow to discharge the capacitor in a short time.

The duration of the discharge, i.e. the time to current zero, is shown by equations (9) and (10) to be proportional to the anode–cathode spacing. Equations (1) and (2) indicate that the theories of Steenbeck and Fünfer predict no such dependence.

An unsatisfactory feature of the above derivation is that at $t=0$ finite currents are obtained, and these may be of considerable magnitude when d is small. This difficulty may be removed by considering the conditions at $t=0$. At this instant the triggering arc is initiated, and it is apparent that the area from which electrons are drawn cannot immediately reach the value A. It may therefore be assumed that expansion of the cross-sectional area of the discharge will take place, i.e.

$$A=\alpha t \qquad \ldots\ldots(11)$$

where α is the rate of expansion.

The current and voltage waveforms can now be calculated as before if A is replaced by the relation of equation (11). Equation (8) thus becomes

$$-C\frac{dV}{dt}=k\alpha t V^{3/2}(d-vt)^{-2}. \qquad \ldots\ldots(12)$$

As a first approximation a linear expansion of area with time will be assumed; α then has a constant value. Together with the appropriate boundary conditions, equation (12) then gives for the voltage

$$V=\left[\frac{k\alpha}{2Cv^2}\left(\frac{vt}{d-vt}-\ln\frac{d}{d-vt}\right)+V_0^{-1/2}\right]^{-2}. \qquad \ldots\ldots(13)$$

Hence the current is

$$i=k\alpha t(d-vt)^{-2}\left[\frac{k\alpha}{2Cv^2}\left(\frac{vt}{d-vt}-\ln\frac{d}{d-vt}\right)+V_0^{-1/2}\right]^{-3}. \qquad \ldots\ldots(14)$$

Equations (13) and (14) are plotted in figure 4 for some representative values, taking as before $k=2{\cdot}34\,\text{A}\,\text{V}^{-3/2}$, $v=10^6\,\text{cm sec}^{-1}$, $V_0=100\,\text{kv}$ and $C=0{\cdot}01\,\mu\text{F}$. The values of α used in the calculations correspond to those estimated in § 3.4.

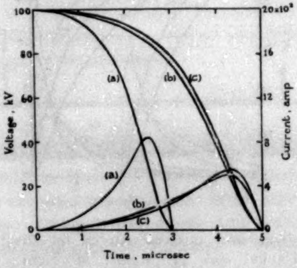


Figure 4. Theoretical voltage and current curves for discharge with expansion. $C=0{\cdot}01\,\mu\text{F}$, $V_0=100\,\text{kv}$. (*a*) $\alpha=7\times10^6\,\text{cm}^2\,\text{sec}^{-1}$, $d=3$ cm. (*b*) $\alpha=7\times10^6\,\text{cm}^2\,\text{sec}^{-1}$, $d=5$ cm. (*c*) $\alpha=6\times10^6\,\text{cm}^2\,\text{sec}^{-1}$, $d=5$ cm.

The equations derived above allow the current and voltage during the discharge to be calculated; (13) and (14) should be used for this purpose unless αt exceeds the total anode area, in which case (9) and (10) should be used.

One of the assumptions made in the above analysis is that the loop inductance of the discharge circuit is sufficiently small for the inductive voltage drop $L\,di/dt$ to be ignored. Examination of the calculated current curves shows that provided the loop inductance is less than $10\,\mu\text{H}$ there will be little error for most of the duration of the discharge. In practice the inductance will normally be between 1 and $10\,\mu\text{H}$. However, the peak current may be subject to a larger error as in this region di/dt can be very high.

§ 3. EXPERIMENTAL RESULTS

Experiments were made with the tube shown in figure 5 for which a tungsten anode and a steel cathode structure were also available. A plane parallel electrode system was used in preference to the normal arrangement with an inclined anode to obtain uniform field and definite anode–cathode spacing. The system was continuously evacuated by an oil diffusion pump and a liquid-air trap was used. The pressure was controlled by an air leak valve, and was measured by a cold-cathode ionization gauge.

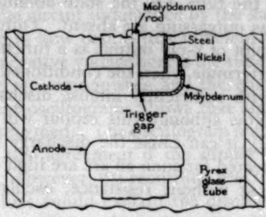


Figure 5. The tube structure.

The electrical circuit is shown in figure 6. All connections were made with thick conductors which were kept as short as possible to obtain a small loop inductance for the discharge circuit. A 20 000 ohm potential divider designed for impulse work was used for voltage measurement. The resistors used for current measurement consisted of flat constantan strip folded over thin mica sheet with the ends close together and soldered to a coaxial connector; the whole was contained in a soft iron can. Measurement showed the inductance of the arrangement to be approximately 5×10^{-8} H. For recording, a high speed cathode-ray oscillograph with a single sweep exponential time base was employed. The signal cable was matched at the receiving end only.

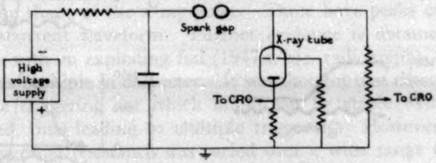


Figure 6. Circuit details.

3.1. *Current and Voltage Waveforms*

Previously published oscillograms of current and voltage waveforms (Slack and Dickson 1947, Fünfer 1950, Reikhrudel, Sanin and Titova 1951) show considerable variation in shape, this being greater for the current waveforms In this investigation waveforms corresponding to all those published have been obtained. It is thus difficult to give examples of 'typical' waveforms as these tend to mislead unless very many are shown.

With a continuously pumped system, the initial behaviour is quite different from the behaviour after a period of use, but even with sealed-off tubes results are not reproducible, as may be seen from the x-ray pulse waveforms obtained by Allen and Hudson (1946).

A selection of oscillograms is shown in figure 7. These all correspond to the conditioned state of the tube, i.e. the state obtained after sufficient use to ensure that effects such as outgassing are absent. Outgassing is associated with low x-ray output, and a plot of x-ray output as a function of the number of discharges can be used to determine when the conditioned state has been achieved.

It is apparent that with sufficiently intense discharges vaporization of the anode will take place. Now although this vapour will not affect the growth of the discharge (see §2.2.2), it can cause the discharge to end as a metallic arc. The conditions to be expected for such an arc are high current, low voltage and, since the circuit is almost without resistance, oscillation of the current and voltage. This is observed experimentally, as shown in figure 7 (*d*) and (*e*). If the discharge does not produce anode vaporization to any large extent then such an arc would not be formed. It would be expected that this case would correspond to a small supply capacitor and/or a large anode–cathode spacing. This again is observed, as may be seen from the aperiodic waveform of figure 7 (*c*). The aperiodic waveforms obtained by Fünfer (1950) were obtained under conditions in which little anode vaporization should occur, viz. a 0·0005 μF capacitor charged to 40 kv. However, using a larger capacitor of 0·01 μF at 40 kv Fünfer did obtain some oscillation. (No figure was given for the anode–cathode spacing, but it is probable that it was small in view of the short discharge time and vaporization at 40 kv. The vaporization at this voltage may be due to the use of a copper anode as it is not clear from the paper which anode material was used.) Further

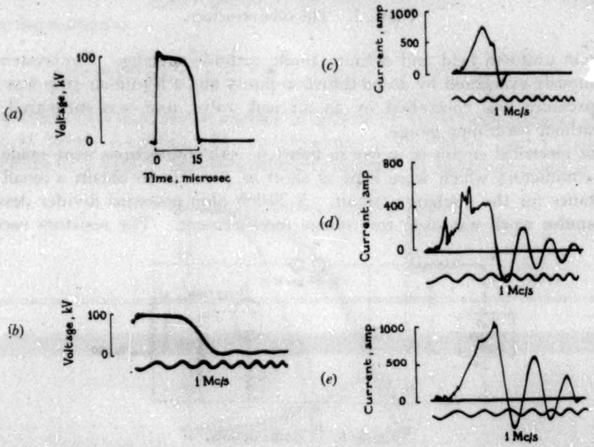


Figure 7. Voltage waveforms $C=0{\cdot}01\ \mu\text{F}$, $V_0=100$ kv. (*a*) $d=14$ cm. (*b*) $d=3{\cdot}5$ cm. Current waveforms $C=0{\cdot}01\ \mu\text{F}$, $V_0=100$ kv. (*c*) $d=4$ cm. (*d*) $d=3{\cdot}5$ cm. (*e*) $d=2$ cm.

confirmation comes from the results of Slack and Dickson (1947) which show the metal arc corresponding to large energy input (0·0067 μF at 300 kv) and small anode–cathode spacing (estimated at approximately 2 cm from photographs of the tube).

Little energy is lost in the arc because although the current is high, the voltage is very low and the circuit resistance extremely small. The voltage causing the arc is that generated in the circuit inductance by the sudden collapse of current. This causes more current to flow than is just sufficient to discharge the capacitor and so produces a small negative voltage across the tube; the resulting oscillation then continues until the arc extinguishes and the energy loss should be small. Integration of the current oscillograms shows that almost all the charge passes before the metallic arc is formed. The oscillogram shown by Slack and Dickson (1947) does not exhibit this agreement if the supply values of the commercial equipment are assumed; however, the recording was made with an experimental circuit, and different values would explain the discrepancy.

The frequency of the oscillation through the metallic arc gives a value for the inductance of the present circuit of 8·5 μH. From the damping of the oscillation the circuit resistance is found to be 4 Ω; of this 0·45 Ω comes from the current measuring resistor and the rest from the resistance of the arc and the radiation resistance since the circuit itself has negligible resistance.

A feature of the current waveforms for which there is no theoretical explanation is the occurrence in some oscillograms of many subsidiary peaks on the rising part of the curve; this is in marked contrast to the theoretical curves, which show a smoothly changing current waveform. These currents do, in fact, flow across the gap and are electron currents since they produce x-rays. This is shown by photomultiplier records (Allen and Hudson 1946, Fünfer 1953) which show the shape of the x-ray pulse. These have peaks corresponding to those on the current waveform. Further evidence is obtained from Clark's measurements with an exploding foil (1947 a, b); radiographs showed that the x-ray burst was multiple in character. It was thought that these peaks might be caused by the triggering arc which might not be stable with the triggering resistance used, thus leading to multiple triggering. However, tests made in which the triggering resistance was varied over a wide range showed that the occurrence of the peaks was not affected. It seems more probable that the peaks are connected with the availability of electrons in the discharge.†

3.2. *The Effect of Pressure Variation within the Tube*

The theories of Steenbeck and Fünfer predict that the peak current occurring in the discharge and the time of discharge will depend on the pressure. As noted previously, Steenbeck obtained agreement between experimental and theoretical results (see figure 2) and found a linear relation between the peak discharge current and pressure, although divergence appeared at the lower pressures. In deriving the alternative theory of § 2.2 it was postulated that the residual pressure within the tube should affect neither the current magnitude nor the discharge time; the following measurements were made to test the validity of this assumption and to find over what pressure range it might apply.

---

† One of the referees has suggested that the effect of the arc in air across the series spark gap should be considered as it may be a factor producing the wave-front oscillations.

The discharge time was selected as a more significant parameter for measurement than the peak current, as its magnitude was less likely to be in error. By discharge time is meant the period elapsing to the first current zero; the oscillatory state of the discharge is, of course, ignored.

A typical set of results is shown in figure 8. It is apparent, despite the scatter of the points, that a change in the magnitude of the discharge duration by a factor of 100, which should occur between 10^{-5} and 10^{-3} mm Hg if the pressure were affecting the results, does not take place. It seems justifiable to assume pressure independence for pressures up to 10^{-3} mm Hg. Unfortunately, at values slightly above this the pressure became uncontrollable, and so no indication is given of the pressure at which residual gas becomes effective. The value obtained here for independence is slightly higher than that obtained by Kingdon and Tanis (1938) for mercury vapour, viz. $0{\cdot}3 \times 10^{-3}$ mm Hg.

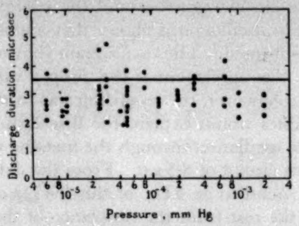


Figure 8. Discharge duration as a function of pressure. $C=0{\cdot}01\ \mu$F, $V_0=100$ kv, $d=3{\cdot}5$ cm. Line shows theoretical value.

Approximate calculations for oxygen show that pressure should become effective for values of pd between about 5×10^{-2} and 2×10^{-3} (where p is the pressure in mm Hg and d is the electrode spacing in cm). Leech (1955) estimates 10^{-1} mm Hg for a 5 cm gap with 50 kv applied. There is thus rough agreement with the experimental values for gaps of centimetre order.

3.3. *The Discharge Duration*

An important result of the theory of § 2.2 is that the duration of the discharge should be directly proportional to the anode–cathode spacing, whilst the theory of Steenbeck and Fünfer predicts no such relation. Thus experimental investigation of the presence or absence of the effect provides a good test of the relative merits of the theories in the high vacuum region.

Some results are shown in figure 9. It can be seen that in general the results tend to confirm the existence of a linear relation between duration and spacing, and that the experimental values tend to lie along the line drawn for a velocity of 10^6 cm sec^{-1} in agreement with the value used in § 2.2.4. This relation has not been noted previously, presumably for two reasons: (i) most other experimental work has been done with x-ray tubes having fixed anode–cathode spacings; (ii) the scatter of the results obscures the effect. At large spacings the discharge duration lies well below the curve around which the values for short spacings are grouped, but it is to be expected that this will occur since both the effect of the tube walls and the effect of residual gas become important for large spacings. The variation in the duration of the discharge at a particular gap setting may be caused by variations in the velocity of the vapour jet from the cathode. Although Tanberg (1930), Kobel (1930), Chiles (1937), and Easton, Lucas and Creedy (1934) all found velocities of the order of 10^6 cm sec^{-1}, differences exist between the values; for several metals the figures given by Chiles are lower than those of Easton, Lucas and Creedy. Also both Tanberg and Kobel found that the measured velocity for a particular experiment varied by a factor of about 3. Further, although the theory of § 2.2.4 was derived on the basis of a single velocity value, it is more likely that a distribution of velocities is present in the jet, and it is to be expected that variations in the energy distribution would lead to variations in the duration of the discharge.

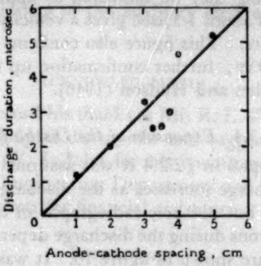


Figure 9. Discharge duration as a function of anode–cathode spacing. $C=0{\cdot}01$ μF, $V_0=100$ kv. Each point is the average of a number of experimental values as shown in the table below :

Anode–cathode spacing (cm)	1·0	2·0	3·0	3·25	3·5	3·75	4·0	5·0
Average discharge duration (μsec)	1·2	2·0	3·3	2·5	2·6	2·9	4·6	5·2
Standard deviation (μsec)	0·1	0·5	1·0	0·6	1·0	0·8	1·6	2·0
Number of values averaged	13	56	49	18	34	28	46	75

It was found that the discharge duration was independent of the voltage applied to the tube over the range 50 to 100 kv for gaps from 1 to 5 cm. This result may be considered in the light of a theory of the production of the high velocity vapour jet (Finkelnburg 1948). Since the jet velocity depends on the vapour production per unit surface area, and therefore on the incident energy per unit area, a change in the velocity should result from a change in the magnitude of the energy delivered per unit area. Now this energy is derived from the anode or cathode falls of the triggering arc, and hence should not be affected by a change in the voltage applied to the tube. A change in the current magnitude may result, but this will affect the velocity only if there is a change in the energy delivered per unit area of surface; this is improbable, as there is evidence that increase of current is compensated either by an increase in the area of the anode and cathode spot, or by the formation of multiple spots (Llewellyn Jones 1953). This probably explains why discharging a large capacitor through the triggering gap, as in the method of Clark (1947 a, b) and Breidenbach (1949), decreased the discharge duration only by a factor of 3 or 4; thus, although it is difficult to increase the energy delivered per unit area, such extreme methods will succeed and result in higher jet velocities and shorter discharge times.

Reference to the values of jet velocity quoted by Easton, Lucas and Creedy (1934) shows that the discharge duration using a molybdenum cathode structure should be shorter than that obtained with a steel cathode structure. Tests were made to detect this difference, but no significant result was obtained, both sets of points tending to lie along a line for a velocity of 10^6 cm sec^{-1}. In view of the variation in values of the jet velocity as discussed above, and the dispersion in the experimental results, it is not thought that this is a serious objection to the proposed mechanism; however, further investigation of this point is required.

A check on the values obtained here is given by the velocity which may be estimated for the tube used by Slack and Dickson (1947) which had a molybdenum cathode structure and was operated at 300 kv. From photographs of the tube the anode–cathode spacing may be estimated as approximately 2 cm and thus the discharge duration of about 1·5 μsec gives a velocity not significantly different from those measured here. This figure also confirms the voltage independence of the velocity up to 300 kv; further confirmation up to 372 kv is obtained from the measurements of Allen and Hudson (1946).

3.4. *Expansion of the Discharge*

In the theory developed in § 2.2.4 it was assumed that the effective area of cross section of the discharge increases as the discharge progresses. This could not be checked directly, but pin-hole radiographs of the anode showed that the area bombarded by electrons during the discharge depended on the anode–cathode spacing. Some results are shown in figure 10. It was found that 50 discharges were necessary to record the full area (no intensifying screens were used). The test using the tungsten anode gives a value for α of 7×10^6 cm^2 sec^{-1} if a value for v of 10^6 cm sec^{-1} is used, and this agrees quite well with the value of about 6 to 7×10^6 cm^2 sec^{-1} obtained from oscillograms of the current during the discharge. The results for the molybdenum anode give a value for α of 3×10^6 cm^2 sec^{-1}; the measured dependence of the recorded area on the number of discharges shows that a factor of two should be used to find the equivalent full area, and the value of 6×10^6 cm^2 sec^{-1} then obtained is in good agreement

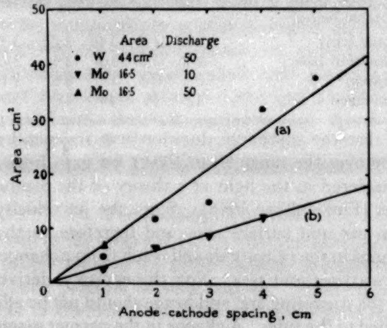


Figure 10. Recorded emitting area as a function of anode–cathode spacing. $C = 0{\cdot}01$ μF, $V_0 = 100$ kv. The curves (*a*) and (*b*) are drawn for values of α of 7×10^6 cm^2 sec^{-1} and 3×10^6 cm^2 sec^{-1} respectively. Inset: Area = anode area; Discharge = number of discharges.

with the values above. The one point obtained for 50 discharges with a molybdenum anode is also plotted in figure 10 and is seen to agree well with the value obtained with the larger tungsten anode.

§ 4. CONCLUSION

The theory proposed here is not at variance with the experimental results; however, it should be regarded as tentative at present as considerably more work is required.

For design purposes two results are of importance: (i) the discharge duration (and hence the x-ray pulse length) depends on the anode–cathode spacing; (ii) attempts to focus by means of the auxiliary electrode will fail since it can exercise no control.

ACKNOWLEDGMENTS

The author wishes to record his thanks to Mr. R. L. Fortescue and Mr. W. K. Mansfield for their many helpful discussions and suggestions during the course of the work, to Professor W. J. John for his encouragement and for facilities provided in the Electrical Engineering Department and to Queen Mary College and the Ministry of Education for financial assistance.

REFERENCES

ALLEN, J. A., and HUDSON, D., 1946, *U.S. Atomic Energy Commission*, MDDC 447.
BRASCH, A., 1947, U.S. Patent 2429217 (October 21).
BRASCH, A., and LANGE, F., 1930, *Naturwissenschaften*, **18**, 16, 765; 1936, U.S. Patent 2043733 (June 9); 1937, U.S. Patent 2099327 (November 16).
BREIDENBACH, H. I., 1949, *Rev. Sci. Instrum.*, **20**, 899.
CHILES, J. A., 1937, *J. Appl. Phys.*, **8**, 622.
CLARK, J. C., 1947 a, *U.S. Atomic Energy Commission* MDDC 768; 1947 b, *Phys. Rev.*, **72**, 741; 1949, *J. Appl. Phys.*, **20**, 363.
CLAYTON, C. G., 1952, *Proc. Instn. Elect. Engrs.*, Pt I, **99**, 206.
CRISCUOLO, E. L., and O'CONNOR, D. T., 1953, *Rev. Sci. Instrum.*, **24**, 944.
DICKSON, D. C., ZAVALES, C. T., and EHRKE, L. F., 1948, *Proc. Nat. Electron. Conf.*, **4**, 298.
EASTON, E. C., LUCAS, F. C., and CREEDY, F., 1934, *Elect. Engng*, **53**, 1454.
FINKELNBURG, W., 1948, *Phys. Rev.*, **74**, 1475.
FÜNFER, E., 1950, *Z. angew. Phys.*, **2**, 25; 1953, *Ibid.*, **5**, 426.
HULL, A. W., and BURGER, E. E., 1928, *Phys. Rev.*, **31**, 1121.
JONES, F. LLEWELLYN, 1953, *Rep. Progr. Phys.*, **16**, 216 (London: Physical Society).
KINGDON, K. H., and TANIS, H. E., 1938, *Phys. Rev.*, **53**, 128.
KOBEL, E., 1930, *Phys. Rev.*, **36**, 1636.
LANGMUIR, I., 1929, *Phys. Rev.*, **33**, 954.
LEECH, J. W., 1955, *Brit. J. Appl. Phys.*, **6**, 107.
LINDELL, B., SIEVERT, R., and WAHLBERG, T., 1950, *Acta Radiol.*, *Stockh.*, **33**, 344.
OOSTERKAMP, W. J., 1940, *Philips Tech. Rev.*, **5**, 22.
REIKHRUDEL, E. M., SANIN, A. A., and TITOVA, T. A., 1951, *Zh. Tekh. Fiz.*, **21**, 746.
SCHAAFS, W., 1949, *Z. angew. Phys.*, **1**, 462.
SCHAAFS, W., and HERRMANN, K. H., 1954, *Z. angew. Phys.*, **6**, 23.
SCHALL, R., 1953, *Arch. tech. Messen*, **208**, 117.
SIEVERT, R. M., 1950, *Acta Radiol.*, *Stockh.*, **33**, 328.
SIMON, L. E., 1947, *German Scientific Establishments* (U.S. Dept. of Commerce, Office of Technical Services) p. 144.
SLACK, C. M., and DICKSON, D. C., 1947, *Proc. Inst. Radio Engrs, N.Y.*, **35**, 600.
SLACK, C. M., and EHRKE, L. F., 1941, *J. Appl. Phys.*, **12**, 165; 1948, *Bull. Amer. Soc. Test. Mat.*, **150**, 1.

SLACK, C. M., EHRKE, L. F., ZAVALES, C. T., and DICKSON, D. C., 1949 a, *J. Soc. Mot. Pict. Engrs.*, **32** (Part 2 : High Speed Photography), 61; 1949 b, *Non-Destr. Test.* **7**, No. 4, 7.
SNODDY, L. B., 1931, *Phys. Rev.*, **37**, 1678.
STEENBECK, M., 1938, *Wiss. Veröff. Siemens*, **17**, 363.
TANBERG, R., 1930, *Phys. Rev.*, **35**, 1080.
THOMER, G., 1953, *Z. angew. Phys.*, **5**, 217.
ZUCKERMAN, V. A., 1943, *C. R. (Doklady) Acad. Sci. U.R.S.S.*, **47**, 267.

RESEARCH NOTES

Note on Turnover in Germanium Contacts

By P. T. LANDSBERG
Marischal College, University of Aberdeen

MS. received 9th January 1956, *and in final form 24th February* 1956

In an important paper, Burgess (1955) showed that many isothermal current–voltage characteristics of germanium diodes, which have been put forward on theoretical grounds, are inadequate in that they yield an incorrect dependence of power dissipation at turnover on ambient temperature. He gives three characteristics which would yield the correct turnover properties. These are, however, of rather unusual form. The problem of determining a more general class of admissible characteristics therefore arises, and is discussed below.

Consider a contact, subject to one-dimensional geometry, carrying a current I and sustaining a voltage V across it. The temperature distribution in the contact along the main axis is assumed to be known in terms of two parameters: the maximum temperature T in the contact, and the ambient temperature T_a. In the steady state the rate of power generation in the contact ($P=IV$) is balanced by the rate R at which energy is lost to the surroundings. It will be assumed that R is proportional to the maximum excess temperature of the contact above the surroundings, so that

$$aIV=T-T_a, \qquad \ldots\ldots(1)$$

where a is a constant. Of the four variables I, V, T, T_a two can be chosen at will, and these will be taken to be the voltage and the ambient temperature. The current and the maximum contact temperature are dependent variables, and their determination requires two equations of which (1) is the first. The second equation will be taken to be an unknown current–voltage relation of the form

$$I=f(V,T,T_a). \qquad \ldots\ldots(2)$$

The maximum temperature of the contact, given V and T_a, is the root of an equation of the form $aVf(V,T,T_a)=T-T_a$. Hence the current can be inferred from (1) or (2), and the temperature distribution can also be calculated.

If the temperature distribution in the contact is uniform, and if $T=T_a$, equation (2) may be regarded as an isothermal current–voltage characteristic. The separate parameter T_a does not then appear in (2). No assumption of this type will be made here. However, any realistic theory of a germanium contact must require (2) to remain a meaningful characteristic in the special case $T=T_a$.

The condition for zero slope in the steady-state characteristic of voltage against current is (Burgess 1955)

$$aV(\partial f/\partial T)_{V,T_a}=1. \qquad \ldots\ldots(3)$$

A subsidiary condition is required to insure that the point of the characteristic located by (3) is in fact a maximum. For any characteristic which describes the experimental findings to a reasonable approximation this condition will be fulfilled, and its existence will not be mentioned explicitly in the sequel. We shall give a suffix t to the values which variables have at turnover, and require, in order to obtain agreement with experiments on germanium, that equations (1) to (3) imply

$$I_\mathrm{t} = I_0, \quad T_\mathrm{t} = T_0, \qquad \ldots\ldots(4a, b)$$

where I_0, T_0 are constants. The results

$$V_\mathrm{t} = (T_0 - T_\mathrm{a})/aI_0, \quad P_\mathrm{t} = (T_0 - T_\mathrm{a})/a \qquad \ldots\ldots(5a, b)$$

are consequences of these conditions. The equations (4) and (5) are experimentally verified approximations for substantial ranges of the variables.

The three isothermal characteristics which were pointed out by Burgess as being in agreement with these requirements are given by

$$I = I_0 F\left(\frac{T_0 - T}{aVI_0}\right), \qquad \ldots\ldots(6)$$

where

$$F(x) = (1 + x)^{-1}, \quad (1 + x/n)^{-n} \text{ or } \mathrm{e}^{-x}.$$

It is very easy to extend this result. For suppose that the isothermal characteristic of the contact has a form (6), where

$$F(0) = 1, \ F'(x) = -1 \text{ if, and only if, } x = 0. \qquad \ldots\ldots(7a, b)$$

Here F' is the differential coefficient of F with respect to its parameter. Now for turnover we have from (3) that $F'[(T_0 - T_\mathrm{t})/aV_\mathrm{t}I_0] = -1$. Hence, by (7*b*), $T_\mathrm{t} = T_0$. It now follows from (6) and (7*a*) that $I_\mathrm{t} = I_0$. We have shown, therefore, that the isothermal characteristics (6), when subject to the conditions (7), lead to the required turnover properties (4) and (5).

It is desirable to enquire if conditions (6) and (7), which have just been shown sufficient for the required turnover properties, are also necessary. We shall show that (6) and (7) are not necessary.

To determine the forms of the characteristic (2), note that (3) implies a relation between V_t, T_t and T_a. There are two cases. In case A, (3) yields the expression (5*a*) for V_t. Since this is valid for all $\mathrm{T_a}$ one can then combine (1), evaluated at turnover, and (5*a*) to find the required properties (4). In case B, the potential cancels out of (3), which then yields a condition whose solution is $T_\mathrm{t} = T_0$. On substituting $T = T_0$ in (2), one must then find $I = I_0$. In this way the correct turnover properties are again obtained.

The result, which is easily verified, is that

$$I = I_0 H(V, T, T_\mathrm{a}), \qquad \ldots\ldots(8)$$

where, case A,

$$H(V_\mathrm{t}, T_0, T_\mathrm{a}) = 1 \text{ and } (\partial H/\partial T)_{V,\,T\mathrm{a}} = (T_0 - T_\mathrm{a})^{-1} \qquad \ldots\ldots(9)$$

at turnover or, case B,

$$H(V, T_0, T_\mathrm{a}) = 1 \text{ and } (\partial H/\partial T)_{V,\,T\mathrm{a}} = (aVI_0)^{-1} \text{ implies } T = T_0. \qquad \ldots\ldots(10)$$

Examples of (9) are

$$H(V, T, T_\mathrm{a}) = \frac{1}{T_0 - T_\mathrm{a}}\,[T - T_\mathrm{a}\,h(V, T_\mathrm{a})], \text{ where } h(V_\mathrm{t}, T_\mathrm{a}) = 1 \qquad \ldots\ldots(11)$$

$$\text{and } \frac{T - T_\mathrm{a}}{T_0 - T_\mathrm{a}} \exp\,[b(V - V_\mathrm{t})], \ b > 0. \qquad \ldots\ldots(12)$$

The first remains meaningful as isothermal characteristic (i.e. when $T_\mathrm{a} = T$), the second does not. This establishes that case A includes characteristics of both types, so that not all characteristics given by (9) are admissible. Examples of (10) are the characteristics given by (6) and (7). To show that (10) defines a more general class than does (6), consider the (isothermal) characteristic

$$H(V,T,T_a)=1+\frac{T_0}{aVI_0}\log\frac{T}{T_0}. \qquad \ldots\ldots(13)$$

It is not of the form (6), but it satisfies (10).

Thus, while the prescription of the correct turnover properties is an important restriction on the isothermal characteristics, it cannot be inferred that the voltage enters into this characteristic only through a parameter of a definite form [such as $(T_0-T)/aVI_0$].

References

Burgess, R. E., 1955, *Proc. Phys. Soc.* B, **68**, 908.

Polarization of the Continuous Spectrum in a Gas Discharge

By B. V. PARANJAPE
Department of Theoretical Physics, University of Liverpool

Communicated by H. Fröhlich; MS. received 19*th December* 1955

Introduction and Discussion

The purpose of the present paper is to show that the continuous part of the spectrum of a gas discharge should be polarized. This is connected with the fact that the velocity distribution of electrons in a gas discharge is not spherically symmetrical. Though collisions between electrons tend to make the velocity distribution random, the external field, through acceleration of electrons, produces a preferred direction of motion.

When a free electron with a kinetic energy E is captured by an ion the electron emits a quantum of frequency $\nu=(E+I)/\hbar$ where I is the ionization energy associated with the energy level into which the electron is captured. The polarization of the emitted photon depends on the direction of observation and on the initial direction of motion of the free electron. (It is a reasonable approximation to assume that the ions are stationary, since the velocities of electrons are much higher than those of the ions.)

To simplify the problem we consider only directions of observation perpendicular to the spark discharge. Let P be the polarization of light near frequency ν due to the recombination of electrons of energy E. It can be written as

$$P(\nu,E)=\frac{I_x(\nu,E)-I_y(\nu,E)}{I_x(\nu,E)+I_y(\nu,E)} \qquad \ldots\ldots(1)$$

where I_x and I_y are respectively the intensities of light polarized along $\mathbf{x}$ and $\mathbf{y}$ axes. Here in a cartesian frame $\mathbf{x}$ and $\mathbf{z}$ are respectively parallel to the direction of the discharge and the direction of observation. The problem therefore reduces essentially to the calculation of I_x and I_y. To avoid complications we restrict the following calculation to the K-shell capture of the electron; thus the detailed results given are applicable only in hydrogen discharge.

Let $i_x(E,\nu,\theta,\phi)$ be the probability that an electron of energy E moving in the direction (θ,ϕ) recombines with an ion, and emits a quantum ν in the direction of observation with its electric vector along the $\mathbf{x}$ direction. The angles θ and ϕ are polar angles, with $\mathbf{z}$ as the axis and ϕ being measured from the $\mathbf{xz}$ plane. Thus

$$I_x = \sum_{\theta\phi} i_x(E,\nu,\theta,\phi) f(E,\theta,\phi) \qquad \ldots\ldots(2)$$

where $f(E,\theta,\phi)$ is the number of electrons having energy E and moving in the direction θ,ϕ.

The probability of a recombination is, by the principle of detailed balance, proportional to its inverse process, the photoelectric effect. Moreover in equation (1) only the relative magnitudes of I_x and I_y are required. Thus for given E and ν (see for example Heitler 1953)

$$i_x(E,\nu,\theta,\phi) \propto e_x(E,\nu,\theta,\phi) \propto \sin^2\theta\cos^2\phi \qquad \ldots\ldots(3)$$

and

$$i_y(E,\nu,\theta,\phi) \propto e_y(E,\nu,\theta,\phi) \propto \sin^2\theta\sin^2\phi \qquad \ldots\ldots(4)$$

where e_x and e_y are respectively the probabilities that a photon of frequency ν polarized in the $\mathbf{x}$ and $\mathbf{y}$ directions releases from an atom an electron of energy E in the direction θ and ϕ.

From equations (1), (2), (3) and (4) it then follows

$$P(\nu,E) = \frac{\sum_{d\Omega} (\cos^2\phi - \sin^2\phi) f(E,\theta,\phi) \sin^2\theta}{\sum_{d\Omega} f(E,\theta,\phi)\sin^2\theta}. \qquad \ldots\ldots(5)$$

As can be seen from equation (5) the polarization depends on the distribution function $f(E,\theta,\phi)$ which has cylindrical symmetry along the $\mathbf{x}$ axis. Taking another system of polar coordinates with $\mathbf{x}$ as polar axis and the angle ϕ_1, being measured from the $\mathbf{xz}$ plane, one can expand the function f in spherical harmonics. Thus

$$f(E,\theta_1) = f_0(E)\mathrm{P}_0 + f_1(E)\mathrm{P}_1(\cos\theta_1) + f_2(E)\mathrm{P}_2(\cos\theta_1) + \ldots. \qquad \ldots\ldots(6)$$

Changing all angles in equation (5) into the above system (indicated by the suffix 1) as

$$\cos\theta = \sin\theta_1\cos\phi_1 \qquad \ldots\ldots(7)$$

$$\cos\phi = \cos\theta_1/[1-\sin^2\theta_1\cos^2\phi_1]^{1/2} \qquad \ldots\ldots(8)$$

we obtain

$$P(\nu,E) = \frac{\int [3\cos^2\theta_1 - 1] f(E,\theta_1)\, d(\cos\theta_1)}{\int [1+\cos^2\theta_1] f(E,\theta_1)\, d(\cos\theta_1)} \qquad \ldots\ldots(9)$$

or using equation (6)

$$P(\nu,E) = \frac{(2/5)f_2(E)}{(4/3)f_0(E) + (2/15)f_2(E)}. \qquad \ldots\ldots(10)$$

CONCLUSIONS

It is well established that the distribution function has the form of a Maxwell distribution, with the electronic temperature of the order 1 ev (see Edels 1950). From ideas developed in the theory of dielectric breakdown in solids (Fröhlich and Paranjape 1956) it can be shown that

$$f(E) = A\exp[-\tfrac{1}{2}m(\mathbf{v}-\mathbf{v}_0)^2/kT] \qquad \ldots\ldots(11)$$

with $E=\frac{1}{2}m\mathbf{v}^2$ where A is a normalization constant, v_0 the average velocity per particle, and T is the electronic temperature. Expanding the left-hand side of the equation (11) in Legendre polynomials as in equation (6) one obtains

$$f_0(E) \propto \sinh b/b \qquad \ldots\ldots(12a)$$

$$f_2(E) \propto 5\,[\sinh b/b - 3\cosh b/b^2 + 3\sinh b/b^3] \qquad \ldots\ldots(12b)$$

where

$$b = mvv_0/kT. \qquad \ldots\ldots(13)$$

Substituting equations (12) in equation (10) one finds the polarization

$$P = \frac{2[\sinh b/b - 3\cosh b/b^2 + 3\sinh b/b^3]}{\frac{4}{3}\sinh b/b + \frac{2}{3}[\sinh b/b - 3\cosh b/b^2 + 3\sinh b/b^3]} \qquad \ldots\ldots(14)$$

It can be seen from equation (14) that the polarization tends to unity as b becomes large, and to zero as b approaches zero. For most discharges the main contribution to the spectral intensity is made by electrons having energies such that $b<1$. This follows from equation (13) since usually $\frac{1}{2}mv_0^2 \ll kT$ and the number of electrons with energy much larger than kT is very small. Thus for $b<1$

$$\begin{aligned} P \simeq b^2/10 &= \tfrac{2}{5}(\tfrac{1}{2}mv^2/kT)(\tfrac{1}{2}mv_0^2/kT) \\ &= \tfrac{2}{5}(\tfrac{1}{2}mv_0^2/kT)(\hbar\nu - I)/kT. \end{aligned} \qquad \ldots\ldots(15)$$

The polarization increases linearly with the energy difference $\hbar\nu - I$. For most electrons $\frac{1}{2}mv^2 \simeq \frac{3}{2}kT$. Thus very roughly the polarization of the continuous spectrum is $\frac{3}{5}(\frac{1}{2}mv_0^2/kT)$. The table gives numerical estimates of polarization for hydrogen discharge in the region of the spectrum that is due to electrons of energy $\frac{3}{2}kT$. The first three columns in the table are taken from Healey and Reed (1941). Here z/p is the ratio of the electric intensity in v cm^{-1} and the pressure of gas in millimetres. $\bar{v}$ is the mean thermal velocity of an electron in cm sec^{-1}.

z/p	20	40	50
$\bar{v}\times 10^{-7}$ (cm sec^{-1})	10·15	13·1	14·0
$v_0\times 10^{-5}$ (cm sec^{-1})	70	160	217
P	0·006	0·02	0·03

Measurement of polarization could thus be used to estimate the electron temperature since it is very simply related to the average drift velocity. Alternatively, from a knowledge of the electronic temperature and the current density one can estimate the density of electrons from the polarization. As can be seen from the table, the polarization increases with the strength of the field and even for slow electrons it is of the order 0·01, which can be measured with reasonable accuracy. It seems possible, however, to create conditions which lead to much higher polarization.

In conclusion it seems that a study of these polarization effects should be of considerable use in the investigation of gas discharge.

ACKNOWLEDGMENT

The author wishes to express his gratitude to Professor H. Fröhlich for help and guidance in all stages of the work. He also wishes to thank Professor J. D. Craggs for discussions regarding experimental aspects of the problem.

The paper is partly based on Report L/T 338 of the British Electrical and Allied Industries Research Association.

REFERENCES

EDELS, H., 1950, *E.R.A. Report* L/T 230.

FRÖHLICH, H., and PARANJAPE, B. V., 1956, *Proc. Phys. Soc.* B, **69**, 21.

HEALEY, R. H., and REED, J. W., 1941, *Behaviour of Slow Electrons in Gases* (Sydney : Amalgamated Wireless of Australia).

HEITLER, W., 1953, *The Quantum Theory of Radiation* (Oxford : University Press).

LETTERS TO THE EDITOR

Electronic Motion in Gases and the Method of Free Paths

§ 1

In a paper " Mecanique statistique des plasmas électronique lorentziens et ses applications à l'ionosphere " by Jancel and Kahan (1955) the authors write (§ 5), referring to the method of free paths, " Huxley a montré que cette méthode conduit à la formule :

$$w_2 = \tfrac{2}{3}(e_2E/m_2)\lambda/\bar{v}_2, \qquad \ldots\ldots (1)$$

pour la vitesse de diffusion en présence d'un champ électrique constant (avec $H_0=0$) ".

It is then implied that I attach great importance to the numerical coefficient $\frac{2}{3}$ and that I am unaware that the value of such a coefficient is determined in part by the law according to which the agitational speeds v_2 of the electrons are distributed. In illustration, they state that according to their treatment, when the law of distribution is that of Maxwell, the coefficient is $8/3\pi$ and not $\frac{2}{3}$.

As the authors seem to have been led astray by formula (1), a brief comment on their statements is called for.

§ 2

It is evident that they have mistaken formula (1), which is certainly incorrect in general, for the following formula for the drift velocity of electrons in gases, that I have discussed and generalized (Huxley 1951) :

$$W = \tfrac{2}{3}(Ee/m)l(\overline{U^{-1}}) \qquad \ldots\ldots (2)$$

where W (their w_2) is the speed of drift, l (their λ) is the mean free path (averaged both with respect to U and to length), and U is the agitational speed of an electron (their v_2).

It will be noted that in formula (1) the drift velocity W (w_2) is proportional to the reciprocal of the mean speed $\bar{v}_2 = \overline{U}$, whereas in formula (2) it is proportional to the mean of the reciprocals of the speeds U. Unless therefore all the speeds U are identical the formulae cannot both be correct; in fact, formula (1) has no theoretical justification.

It is, however, a simple matter to introduce the mean speed $\overline{U}$ into formula (2) as follows :

$$W = \tfrac{2}{3}(\overline{U}\,.\,\overline{U^{-1}})(Ee/m)l/\overline{U}. \qquad \ldots\ldots (3)$$

The numerical coefficient in formula (3) is the product of the factor $\frac{2}{3}$ and a dimensionless factor $\overline{U}\,\overline{U^{-1}}$. The value of the former derives from the assumption of chaotic agitational motion in the group of electrons, whereas that of the latter depends upon the law assumed for the distribution of the speeds U; consequently a modification of either assumption produces in general a change in the value of the numerical coefficient in formula (3). If Maxwell's law of distribution be adopted then $\overline{U}\,.\,\overline{U^{-1}} = 4/\pi$ and $\frac{2}{3}(\overline{U}\,.\,\overline{U^{-1}}) = 8/3\pi$ which is the value found by Jancel and Kahan. This technique of using dimensionless factors to take account of the effect of the distribution of the speeds U upon numerical factors in formulae relating to the motions of electrons in gases has been extensively used for some time (Huxley 1940, 1951, Huxley and Zaazou 1949, Crompton and Sutton 1952, Crompton, Huxley and Sutton 1953) and it is therefore not clear what is the basis of the comment by Jancel and Kahan that " cet auteur semble attacher une grande importance au facteur 2/3 bien qu'il ne précise pas la nature de la fonction de distribution qui determine v_2 ".

§ 3

Jancel and Kahan write : " En raison de sa nature même la méthode du libre parcours moyen presente l'inconvénient de ne s'appliquer qu'à des champs faibles, sans qu'il soit possible de préciser d'une manière rigoreuse la validité et l'ordre de grandeur des approximations. De plus, le fait de conduire le calcul en supposant d'abord que tous les électrons ont la même vitesse après le choc entraine des divergences importants . . . ". As these statements are factually no more satisfactory than those discussed above it is useful to consider the physical model upon which formula (2) rests. It is supposed that the paths of electrons among the gas molecules (when $E=0$) are rectilinear segments such that the change in direction from each segment to its successor occurs in a distance small compared with the mean length of the rectilinear segments. The paths are traversed at speeds U which are distributed according to some law, but there is no restriction of $\overline{U}$ to a value, as implied by Jancel and Kahan, corresponding to equipartition in extremely weak fields.

When the scattering is isotropic the drift velocity of the centroid of those electrons with agitational speeds U is $W_{\mathrm{U}}=\frac{2}{3}(Ee/m)(l/U)$. Since the mass of an electron is very small compared with that of a molecule the proportion of its energy lost by an electron in a single collision is, on the average, small. Consequently, a group with speed U retains its identity while its members individually make a number of collisions with the molecules.

The mean drift velocity of the whole assemblage is therefore

$$W=\overline{W_{\mathrm{U}}}=\tfrac{2}{3}(Ee/m)(\overline{l/U}). \qquad \ldots\ldots (4)$$

Although equation (4) is exact, nothing is known *a priori* about the dependence of the mean free path upon the speed of the electron. However, if it is assumed to be independent of the speed U then the general formula (4) reduces to the restricted formula (3).

The fact that both formula (3) and the general expression of Jancel and Kahan lead to the value $8/3\pi$ for the coefficient in formula (3) when Maxwell's law of distribution is postulated is not a coincidence but merely an expression of the fact that each deduction in this special instance is employing, in effect, the same physical model.

Since it is explained elsewhere (Appendix to Hall 1955) how this model of electronic motion can be made to conform with the macroscopic properties of electronic motion in actual gases, it is not necessary to discuss this matter further.

The poor esteem in which the method of free paths is held derives from its inadequacy in the kinetic theory of gases, but when applied to the theory of motion of electrons in gases it is in some respects, because of the simplifying assumptions that are legitimate, but little inferior to the Maxwell–Boltzmann procedure and usually more convenient.

It may be remarked that the theory of the motion of electrons in gases in the presence of steady electric and magnetic fields was thoroughly investigated many years ago by Pidduck (1915, 1936) but his work appears to have been overlooked in the majority of recent investigations that are based on the Maxwell–Boltzmann procedure.

Physics Department, L. G. H. Huxley.
University of Adelaide.

6th July 1955, in final form 20th February 1956.

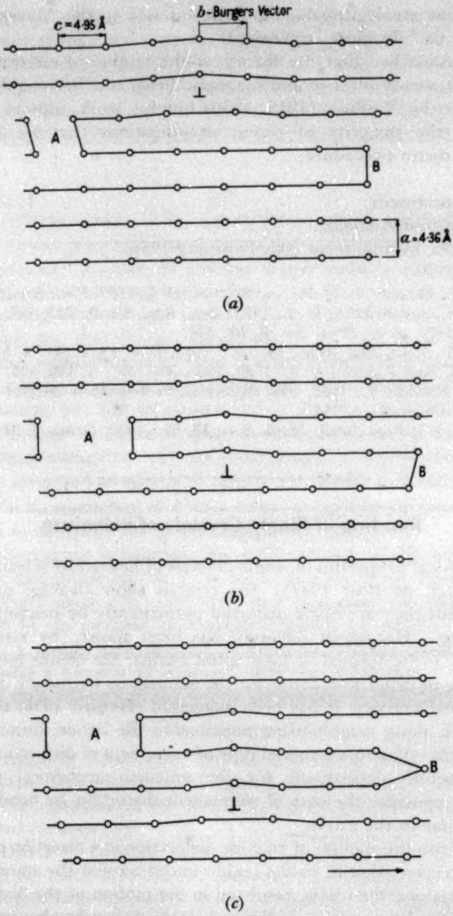
Crompton, R. W., Huxley, L. G. H., and Sutton, D. J., 1953, *Proc. Roy. Soc.* A, **218**, 507.
Crompton, R. W., and Sutton, D. J., 1952, *Proc, Roy. Soc.* A, **215**, 467.
Hall, B. I. H., 1955, *Proc. Phys. Soc.* B, **68**, 334.
Huxley, L. G. H., 1940, *Phil. Mag.*, **29**, 327; 1951, *Proc. Phys. Soc.* B, **54**, 844.
Huxley, L. G. H., and Zaazou, A. A., 1949, *Proc. Roy. Soc.* A, **196**, 402.
Jancel, R., and Kahan, T., 1955, *The Physics of the Ionosphere* (Report of the Physical Society Conference), p. 365.
Pidduck, F. B., 1915, *Proc. Lond. Math. Soc.*, **15**, 89; 1936, *Quart. J. Math.*, **7**, 27, 189.

Bending of Single Crystals of Selenium

The mechanical properties of single crystals of hexagonal selenium are somewhat unusual (cf. de Boer 1947) : the crystals show cleavage only along the prism planes, but they are easily distorted permanently by bending, twisting or simply crushing. Hexagonal selenium has been shown, by x-ray diffraction (Bradley 1924), to be formed of long spiral chains, the chains lying parallel to the *c* axis; and, in order to account for the anomalous frequency dependence of the electrical conductivity, it has been suggested (Hyman 1956) that the spiral chains fold back along neighbouring positions in the lattice, forming a' zigzag ' structure. In this structure a special type of movement of dislocations is possible which can account qualitatively for the unusual mechanical properties of selenium. We consider the case of permanent distortion by bending about an axis perpendicular to the *c* axis.

Figure 1 shows the motion of an edge dislocation in a direction perpendicular to the Burgers vector, that is, in the region under tension the spiral chains slide over each other along the *c* axis, resulting in the motion of the dislocation away from the surface under tension. In figure 1 (*b*) the dislocation has moved through the two lattice spacings, and in 1 (*c*) through a further lattice spacing. As the dislocation moves in this manner, it forms a series of holes, such as the holes formed at points A and B.

In order to account for deformation by means of dislocations, it is necessary to have a source of dislocations. In the present case the dislocations can simply be formed at the surface, because the zigzag structure is expected to extend to the surface and there is only an adsorbed layer of gas on the surface. The dislocations can then move, in the manner described above, to positions corresponding to the degree of bending.



Schematic drawing of movement of edge dislocation in hexagonal selenium. Dislocation line perpendicular to *c* axis and at 60° into paper.

Standard Telecommunication Laboratories Ltd.,
Enfield,
Middx.
22nd March 1956.

R. A. HYMAN.

DE BOER, F., 1947, *Philips Res. Rep.*, **2,** 349.
BRADLEY, A. J., 1924, *Phil. Mag.*, **48,** 447.
HYMAN, R. A., 1956, *Proc. Phys. Soc.* B, **69,** 743.

REVIEWS OF BOOKS

Linear Feedback Analysis, by J. G. THOMASON. Pp. x + 355. (London: Pergamon Press, 1955.) 55*s*.

In the last 20 years, following the fundamental work of Nyquist and of Bode, feedback theory has not only become an indispensable tool of the engineer, but by its ramifications is likely to affect such sciences as biology and economics. Can this book be recommended to the physicist who would like to know " what it is all about ", or who would like to understand the feedback instruments of which he makes daily use ? The answer to the first question is " no ", to the second a qualified " yes ".

The author intends his book " to provide an analytical background for the young science graduate working in electronics or servomechanism research ". It is somewhat surprising that he finds it necessary to teach this analytical background starting from a level at which not even complex numbers or the solution of linear equations can be assumed as known. Granting that the author bases his views on practical experience, this is a crushing reflection indeed on the type of science graduate who takes up electronics. It appears doubtful whether this type will ever acquire other than an instinctive and purely practical knowledge of the subject, in spite of the great trouble which the author takes introducing every bit of mathematics in closest connection to its practical applications. It must be regretted that, having in mind mainly this kind of trainee, the author has just missed making his book a good introduction to people who really deserve to be called ' science graduates '. A research worker with even the rudiments of a scientific outlook can hardly be satisfied with a book which gives the formula for the (direct) Laplace transform, but dismisses the inverse transform with the remark that " it is not defined so simply, but comprehensive lists of transforms are available " (p. 29).

Another weakness of the book is one common to many experts who write on their own field; the nearer they come to their ' daily bread ' the less they succeed in making it clear to those to whom it is new. Nyquist diagrams are the daily bread of the servo-engineer, and Section 6.7 which deals with these is in parts almost unintelligible. For instance the eq. 79 on p. 152, $C = \cos\phi$, " between gain or modulus C and phase shift " will completely puzzle the reader, and he will not much profit from the reference to Section 3.1, as in that section C always stands for a capacitance. Even if the misprint is corrected, and the equation written $C = L\cos\phi$, it will still puzzle the reader unless he remembers that from line 9 on p. 152 the symbol L has come to mean not $L(p)$, but $L(0)$, the zero frequency loop gain. To make matters worse, in the explanatory figure 6.36 the L-plane is labelled " p-plane ". Such small errors can be of course easily overlooked by the expert author, but they can drive the beginner to despair.

The centre-piece of this book is Hendrik Bode's theorem on the relation between gain and phase shift, and here the author has failed to explain why a relation of this sort must exist, that is to say why a complex function which has no poles and zeros at one side of an axis is fully determined by its real (or imaginary) values on that axis. It is unlikely that the reader will find this out for himself, oppressed as he will be by the mass of rather complicated and for him novel mathematics, and he will probably accept Bode's equation as a dogmatic rule, which must be believed and memorized.

These, and other similar shortcomings, which could have been easily avoided, are regrettable in a book of high practical merit and usefulness, and it is hoped that they will be remedied in the second edition, which it is likely to attain owing to the excellence of its practical chapters.

The contents are: 1, Steady-state circuit analysis; 2, The Laplace transformation; 3, Illustration of the transformation technique; 4, Introduction to feedback circuits; 5, Amplifying stage design; 6, The stability of feedback systems; 7, Gain–phase analysis; 8, Stabilization techniques; 9, Illustrative feedback amplifiers; 10, Feedback integrators and differentiators; 11, Stabilized power supplies, with useful lists of Laplace transforms and of valve equivalents.

D. GABOR.

Progress in Low Temperature Physics, edited by C. J. GORTER. Pp. xii + 418. (Amsterdam: North-Holland Publishing Co., 1955.) 61*s*.

Among the subjects dealt with in the eighteen articles in this volume are the two-fluid model for superconductors (Gorter), the application of quantum mechanics to liquid helium (Feynman), heat conduction in superconductors (Mendelssohn), the electronic specific heat in metals (Daunt), adiabatic demagnetization (de Klerk and Steenland), and transport phenomena in gases at low temperatures (de Boer). There are also four articles on paramagnetism, ferromagnetism and anti-ferromagnetism, and three other articles on liquid helium (including liquid ^{3}He). As must be expected in a combined effort of this kind, the standard varies from the excellent downwards. To a few of the authors the editor seems to have been rather lenient.

There is no doubt that this, and the succeeding volume which is expected, will prove invaluable to low-temperature physicists. If the second volume is of the same standard as the first they will give a definitive and fairly detailed picture of the present state of the subject of a kind which does not yet exist. The reports of international conferences—which have appeared quite frequently in recent years—are incomplete in the same way as are supplementary Meccano "A" sets without the sets which they supplement. It is however something of a relief to find that it is apparently not the intention to make these into annual progress reports. Experience in low-temperature conferences suggests that the subject—and authors—could hardly sustain this at present. G. O. J.

Chemistry of the Solid State, edited by W. E. GARNER. Pp. viii + 417. (London: Butterworths Scientific Publications, 1955.) 50*s*.

The physics and chemistry of the solid state have received considerable attention from physical chemists since the pioneer work of Garner, Mott, Schottky and others. The physical aspects of this subject have of course been treated in a number of books such as *Dislocations and Plastic Flow in Crystals*, by Cottrell, *Electronic Processes in Ionic Crystals*, by Mott and Gurney, and *Modern Theory of Solids*, by Seitz. Professor Garner has recently edited a book written by specialists, which for the first time brings together published information on the chemical aspects, in a form convenient to those actively engaged in this field.

This book falls into two parts. The first, consisting of seven chapters, presents the more theoretical developments and deals with the following topics: 1. The chemistry of crystal dislocations; 2. Lattice defects in ionic crystals; 3. Action of light on solids; 4. Surfaces of solids; 5. Semi-conductivity and magneto-chemistry of the solid state; 6. Theory of crystal nucleation from vapour, liquid and solid; 7. Classification and theory of solid reactions. The second part is mainly concerned with the experimental results and the application of the principles described in the first part. The subjects dealt with are: 8. Kinetics of endothermic solid reactions; 9. Kinetics of exothermic solid reactions; 10. Decomposition of organic solids; 11. Explosion and detonation in solids; 12. Solid–solid reactions; 13. Photographic process; 14. Oxidation of metals; 15. Electronic factor in chemisorption and catalysis.

It is impossible in the space available to comment adequately on each section. The book as a whole has been written in a clear and concise manner, and this reviewer found particular pleasure in reading chapters 3 and 7 by Jacobs and Tompkins. In the sections describing the thermal decomposition of solids it is evident that we have no clear understanding of the basic mechanism of the decomposition process. The decompositions have been followed in the main by plotting pressure–time curves, and an analysis of such curves does help in classifying the different kinds of reactions. However, in order to get at the detailed mechanism it will probably be necessary to extend the use of such techniques as x-ray diffraction, electron diffraction, electron microscopy and reflection electron microscopy. Chapter 13 on the photographic process by Mitchell is an excellent example of the way in which ideas and theories developed in solid state physics are applied to a specific problem. It is inevitable in a publication of this kind that there should be some lack of coherence between the various chapters. There are also a number of noticeable gaps. One such omission is the recent work of Eggert and his colleagues on the decomposition and ignition of endothermic compounds by light. Another is the effect of ionizing radiation and nuclear particles on the properties of crystalline solids and plastics.

Although it is stated in the preface that this book was written primarily for physical chemists, it should in fact have a much wider appeal and may be read with profit by all scientists working in related fields. There is no doubt that it will be the standard reference work for a number of years and it should suggest and stimulate further research. A. D. YOFFE.

Luminescence and the Scintillation Counter, by S. C. CURRAN. Pp. x+219. (London: Butterworths Scientific Publications, 1953.) 32*s*. 6*d*.

Perhaps in sympathy with the teacher's aphorism 'those that can, do', a readier ear is accorded the creator–expositor; for a subject in the hot surge of advance, there is no better guide. This little book, in twelve short chapters, is the work of one such: for the scintillation counter in its modern form was first developed by Dr. Curran but ten years ago. Already it has proved itself, thanks to its speed, sensitivity, flexibility, and discriminating power (in conjunction with associated circuitry and accessories) as one of the chief tools in the quantitative analysis of particles and radiations of all types, particularly those of nuclear origin. The present account, more specialized than the cognate *Counting Tubes* and designed for serious workers aiming to develop or use the technique, is the first integrated summary of all aspects.

Following a brief introduction, two chapters indicate the essentials of radiation–matter interaction, with special emphasis on secondary emission. Mesons and neutrinos are included. A direct lead to multipliers is given in consideration of high secondaries yield, lattice potential and energy factors favouring the higher alkali halide complexes. Practical detail follows in the next two chapters describing general principles of the electron multiplier and characteristics of some commercial types, chiefly the ' venetian blind ' and shaped dynode models, the former, though less subject to magnetic disturbance, being slower in response and with less powerfully defined field control. Features necessary in the photosensitive surfaces are fully explained and practical materials given. Figures of merit are based on spectral sensitivity, dark current, noise–temperature relation, amplification, intensity range, stages, shielding, collection efficiency, and time and amplitude resolution. Difficulties such as satellite pulses are not ignored.

A similar treatment of luminescence follows, two chapters sketching current views on fluorescence and phosphorescence, followed by two more on the preparation and properties of practical scintillants. Clearly the author has critical reservations here on established concepts and general notions, but indicates the importance of ring and bond resonance, high melting points and absence of chromophores in scintillants, of matching spectral sensitivities of scintillant and multiplier, and indicates how modern excitation methods have weeded out theories of energy transport, if only on the basis of decay times. The practical chapters detail chemical range, purification processes, phases (including plastic matrices and polymerization *in situ*) and give ample coverage to nature and properties with a natural bias to the organic crystal and solution scintillants whose extreme speed of response serves in analysis and applications.

The two streams reunite in the final three chapters, the first and last of which are especially valuable for critical assessment of suitable applications and an excellent summary of circuitry for purposes including time resolution, fast scaling, display, discrimination, pulse-sorting, or coincidence work, with frequent illustration from the Glasgow and other laboratories.

Designed for use, first rather than last words, a bench handbook rather than a sofa volume, for Ramsays rather than Rayleighs, in style compact, terse, staccato, selective, the book is suggestive and compelling with few irrelevancies, though an occasional ' salute to adventurers ' gleams tribute to solid or deft analysis or technical virtuosity. Though coverage is very full and the 340 references, international in scope, include some books and reviews, in the inevitable revision further developments in medical and low-energy x-ray applications might find a place. There are a few omissions, particularly in the subject index. It must be a publisher's irony that a scientific research book's scintillating success helps to decrease its decay time ; but already it seems a far cry to those Royal Institution lectures where Rutherford, with the aid of Crowe, Green and Wynn-Williams, described the then latest ' lazy methods '.

I. MACARTHUR.

Thermodynamics and Statistical Mechanics, by W. P. Allis and M. A. Herlin. Pp. viii + 239. (New York, Toronto, London: McGraw-Hill, 1953.) 51*s*.

This book was written as a text for seniors in physics at the Massachusetts Institute of Technology, and the authors express their indebtedness to Zemansky's *Heat and Thermodynamics* and Slater's *Introduction to Chemical Physics* in the preparation of it. While it clearly owes something to both these sources it has at the same time a pleasing character of its own, and the reviewer has read it with enjoyment and, he believes, profit. As a text for honours physics students in this country it would seem, however, to suffer from the drawback of not going quite sufficiently far in either subject.

The thermodynamic half of the book is rather informally presented and, perhaps for this reason, not particularly easy reading. Its chief weakness is in its treatment of temperature: only the ideal-gas scale is mentioned and the identification of this with the Kelvin scale is rather cavalier. On the other hand its emphasis on specific heats and their dependence on experimental conditions is to be welcomed. Discussion is confined throughout to one-component systems and first-order phase changes.

The introduction to statistical mechanics, which forms the second half of the book, is based on the Boltzmann–Planck foundation of cells in phase-space. Discussion is almost entirely confined to perfect gases, though the effects of gravitational and magnetic external fields are considered and degenerate gases, obeying Einstein–Bose or Fermi–Dirac statistics, are dealt with (rather briefly). A brief discussion of thermal noise (Nyquist's theorem) precedes the treatment of radiation laws and is an unusual and attractive feature of the presentation. Perhaps its chief merit, however, is that Stirling's formula is never used except for numbers which really are large. In this the treatment is very elegant. On the other hand it is a pity that quantum mechanics is made to appear so mysterious, and that there is not even an adequate discussion of so simple a matter as the specific heats of diatomic gases. This seems a heavy price to pay for the decision not to introduce the concept of partition function.

Each half of the book should be read straight through; it is not a book for dipping into. There is an excellent collection of relatively simple examples at the end of each chapter (except the last, on degenerate gases).

G.S.R.

Nuclear Magnetic Resonance, by E. R. Andrew. Pp. xi + 265. (Cambridge: University Press, 1955.) 35*s*.

Although it is now a little over ten years since the birth of nuclear magnetic resonance, no treatise on the subject has appeared apart from the valuable thesis of Bloembergen (1948), which was rapidly sold out and has since been out of print. An authoritative treatise is still lacking, but this volume by Professor E. R. Andrew, the latest in the series of Cambridge Monographs on Physics, gives as comprehensive a review as its size permits. The basic theory cannot be developed in full on such a scale, and the author has rightly relegated some of the more tedious mathematics to appendices, or given useful order of magnitude arguments, for example in the subject of spin–lattice relaxation times. These make for easy reading for the beginner, who can thus concentrate on the physics; but it seems a pity that more use is not made of rotating coordinate systems, which are very valuable in forming a physical picture of the phenomena, particularly when dealing with transient effects.

After a brief introduction follow chapters on the basic theory and experimental methods, the latter including references to methods of optical detection and nuclear alignment. The measurement of nuclear moments and applications to some experiments in fundamental physics are then discussed, followed by three chapters on the information which nuclear magnetic resonance has given on effects in bulk material arising from the interaction of the nuclear moments with their surroundings. These will be of particular value to the applied physicist. The book ends with a short chapter on electric quadrupole effects, together with a valuable bibliography and useful tables of substances investigated, including references. A singular omission is of any table of nuclear spins and magnetic moments; the latter is excused on the grounds that moments are not of great value unless details of corrections are given. But surely, now that the spins of practically all stable isotopes have been determined, and their moments are known with fair accuracy, these could have been included in a handy table together with the references to the original experiments.

B. BLEANEY.

Niels Bohr and the Development of Physics, edited by W. PAULI. Pp. vii + 195. (London: Pergamon Press, 1955.) 30*s*.

This little book has been prepared as a tribute to Professor Bohr on his seventieth birthday. The authors of its ten chapters are distinguished physicists who have collaborated with Professor Bohr in the many branches of physics in which he has played such a central role. In a book of this kind, one hopes to read of things which will endear the great man to those who have not had the privilege and inspiration of working with him in person. Such articles there are; but some there are too which seem hardly appropriate to the occasion. Nevertheless, after reading this collection of papers one is left primarily with a feeling of wonder that one man should have done so much.

B. H. FLOWERS.

Elektrische Lichtquellen : Gasentladungslampen (Electric Light Sources: Gas-discharge Lamps), by A. P. IWANOW. Pp. xii + 583. (Berlin: Akademie-Verlag, 1955.) 55 DM.

This is a recent German translation of a volume originally published in Russian in 1948. It summarizes current theory and practice in the U.S.S.R., and part of the contents will assuredly be of considerable interest to those physicists and electro-technicians concerned with the production and exploitation of electric light sources.

The text consists of 23 chapters divided into eight sections which are (1) an elementary introduction to atomic spectra, (2) the basic collision phenomena in discharge tubes, (3) properties of glow discharges, (4) glow and arc discharge lamps, (5) metal vapour tubes, including high pressure mercury lamps, (6) discharge tubes with light yield heightened by fluorescence, (7) the arc lamp, (8) miscellaneous sources.

Since the original volume was intended as a text for technical students in the U.S.S.R. there is a particularly marked concentration on the techniques in vogue in that country and also on references in Russian. Indeed this constitutes a particular value of this German text for those of us who cannot read Russian. Of a total bibliography of 268 titles no less than 93 refer to Russian publications in U.S.S.R. journals. A peculiar omission is the complete absence of an index, although it is true that the contents page has 100 sub-titles.

The emphasis is primarily concerned with light sources of technological and industrial importance and the many sources of considerable academic and purely scientific interest are hardly surveyed at all. For example, the spectroscopist and the user of interferometric devices will hardly find anything here about the particular electric light sources which would interest him, so that of course the formidable literature on such spectroscopic and optical sources is not considered. Nevertheless the book contains a great deal of very detailed practical information, especially in relation to industrial types of light sources developed specifically in the U.S.S.R. This detail is everywhere backed by comprehensive theoretical treatment where such treatment is known.

Sufficient detail is given of the many novel products of the Moscow Electro-Lamp Factory such that a translation into English might well prove beneficial to the lamp industries of Great Britain and the U.S.A.

In introductory sections to some of the chapters the author expounds some individual views on the early history of several long established lighting devices, offering documentary support for his conclusions. S.T.

Fundamentals of Radiobiology, by Z. M. BACQ and P. ALEXANDER. Pp. xii + 389. (London: Butterworths Scientific Publications, 1955.) 40*s*.

This volume provides a fascinating and readable account of the more significant aspects of radiobiology at the present time and is written primarily for the non-specialist. Although the physicist would find the introductory chapter on the effects of ionizing radiations on matter somewhat superficial, the descriptions of some of the complex biochemical and physiological aspects of the subject in the later chapters may be read by physicists with very great profit. There is an excellent index extending over 30 pages and 960 original papers are quoted in the bibliographies. A more detailed review of this volume will be found in Part A of *The Proceedings* (p. 500). W. G. MARLEY.

Humidity, by H. L. PENMAN. Pp. 72. (London: Institute of Physics Monographs for Students, 1955.) 5*s*.

This series of monographs is intended for general reading in courses for the Higher National Certificate in Applied Physics and the first two years of degree courses. In this small volume Dr. Penman discusses the physical principles underlying the measurement of humidity and briefly describes the construction, operation and use of various types of hygrometer. Here is a clear, straightforward treatment of a subject of considerable practical importance. Although the author's derivations of the psychrometer equation and the expression for the equilibrium vapour pressure over a curved surface may lack something in rigour and elegance, they are probably adequate for his purpose.

One might, perhaps, remark that the statement, made on page 11, that "The equilibrium partial pressure of water vapour at a given temperature is the same whether there are other gases present or not", is not quite correct, and suggest that the author's *cri de cœur*, "To give the temperature and relative humidity of the air is good physics; to give only the relative humidity is meaningless jargon", rather overstates a case which might have been presented more convincingly.

But these are minor points. This book will undoubtedly prove very useful to anyone confronted with the practical problem of measuring humidity and can be recommended. B. J. MASON.

CONTENTS OF SECTION A

PAGE

Mr. V. HEINE. The de Haas–van Alphen Effect in Alloys 505

Mr. V. HEINE. The Band Structure of Bismuth 513

Mlle. M. CHARTON and Dr. A. G. GAYDON. Band Spectra Emitted by Strontium and Barium in Arcs and Flames 520

Dr. R. H. DALITZ. Isotopic Spin Changes in τ and θ Decay 527

Mr. P. S. FISHER and Mr. V. KNAPP. The Decay of the Isomeric State of ^{198}Tl . 541

Dr. D. P. JENKINS. Calculations on the Band Structure of Silicon . . . 548

Mr. J. E. BENNETT, Dr. D. J. E. INGRAM and Dr. D. SCHONLAND. Paramagnetic Resonance Spectra of the Alkali Superoxides and Chlorine Dioxide . . 556

Dr. E. BREITENBERGER. Attenuated Angular Correlations between K-conversion Electrons in ^{197}Hg 566

Research Notes:

Dr. R. F. BARROW, Mr. W. B. GRATZER and Mr. J. F. MALHERBE. Intensities of Bands in the System $b^3\Sigma^+ - a^3\Pi$ of Carbon Monoxide 574

Reviews of Books 577

Contents of Section B 580

Avalanche Injection in Semiconductors

By J. B. GUNN
Radar Research Establishment, Great Malvern, Worcs.

MS. received 18*th January* 1956

Abstract. It is predicted on theoretical grounds that additional charge carriers may be injected into a semiconductor by a thin region of high electric field in which avalanche multiplication occurs. The voltage drop across the region may be quite low and associated with a negative resistance for certain values of the current. The observation of the predicted phenomena in germanium is described; it is found that the negative resistance region may cover some hundreds of volts and the response time be as short as 3×10^{-9} sec. The implications of avalanche injection are discussed in connection with solid-state phenomena involving a large current density, and it is shown that any sufficiently small contact is capable of rectification independently of the presence of potential barriers.

---

§ 1. Introduction

The effects of a large electric field on the holes and electrons of an elementary semiconductor have been the subject of recent theoretical calculations by Shockley (1951), Wolff (1954) and by Yamashita and Watanabe (1954), and of experimental measurements by Ryder (1953), McKay (1954), McKay and McAfee (1953), Miller (1955) and by the author (1956). The results obtained for germanium may be summarized as follows. At fields less than about 10^3 v cm^{-1}, the average drift velocity of the carriers increases approximately proportionately to the electric field, and the carrier 'temperature' remains in equilibrium with that of the lattice. At larger fields, a departure from this linear relation occurs and the drift velocity rises less rapidly than the electric field, becoming about equal for holes and electrons and almost independent of the value of the field over the range $3 \times 10^3 - 6 \times 10^4$ v cm^{-1} (Gunn 1956). In this condition the carrier temperature becomes much larger than that of the lattice. At the upper limit of the range the electrons are able to acquire enough energy to ionize some of the valence bonds of the lattice and to liberate a mobile carrier pair. Since, as shown by McKay, both of the new carriers can cause further ionization, the process, termed avalanche ionization, is a cumulative one and may lead to a very rapid increase in the carrier density. The hole and electron currents in an avalanche have each a finite divergence, and this, together with the saturated drift velocity, gives rise to a space charge, and hence to a distortion of the electric field.

In the light of these results we may consider the phenomena that may be expected to occur in a system such as that shown in figure 1 (*a*) when the current density is raised to large values. Here carrier injection and potential drops at the electrodes are both supposed to be prevented by the nature of the electrodes. Suppose first that the central region is extrinsic n-type with equilibrium electron density n_0. Then, as the current is increased from zero, the voltage will at first rise proportionately, as at OA in figure 1 (*b*). When current densities of the order of $J_0 = qn_0v_s$ (v_s = saturated drift velocity) are reached, the voltage will rise extremely rapidly, while the carrier density remains equal to n_0 and the field remains uniform. When fields as high as $6 \times 10^4\,\mathrm{v\,cm^{-1}}$ are reached avalanche ionization will begin, and the additional carriers so produced will slow down the rise of voltage and of mean electric field (point B, figure 1 (*b*)); their space charge will, however, lead to a field distortion which results in the field near the positive electrode being considerably larger than the mean field, while the field near the negative electrode becomes less. Further increases in current will lead eventually to a situation in which the field at the negative electrode is reduced to a low value (point C), and at currents beyond this the potential drop becomes increasingly confined to a thin region of very high electric field near the positive electrode, the field in the remainder of the specimen being reduced almost to zero. The thin avalanche region is, however, a prolific source of holes, which are injected into the low field region, so increasing its conductivity and enabling the large current to be conducted in the presence of only a small electric field. The term *avalanche injection* is proposed as a description of this phenomenon.

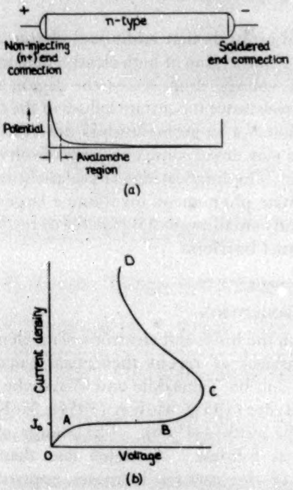


Figure 1 (*a*). The simplest form of injecting avalanche. (*b*) Expected form of current–voltage relationship.

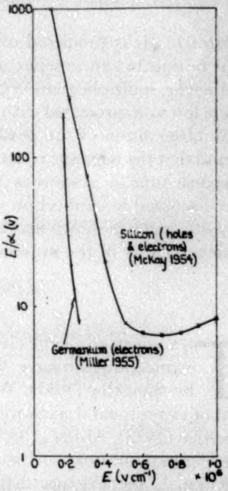


Figure 2. Variation of E/α with field.

It is readily shown that, over at least the first part of the region CD of figure 1 (*b*), the voltage must be decreasing with increasing current. For suppose the current to be large compared with J_0, so that the multiplication M occurring in the avalanche region is large. Then, if α is the avalanche ionization coefficient of McKay (1954), it is known that $1 - 1/M$ is of the order of

$$\int_0^{x_a} \alpha\,dx,$$

where x_a is the width of the avalanche region (the exact value depends on the boundary conditions in the avalanche region). Similarly, the voltage V_a across the avalanche region due to the field E within it is

$$-V_a = \int_0^{x_a} E\,dx.$$

To the degree of approximation required here the integrals may be replaced by the products of x_a with the mean values $\bar{\alpha}$ and $\bar{E}$, and $1/M$ may be neglected by comparison with unity. Hence $-V_a \sim \bar{E}x_a \sim \bar{E}/\bar{\alpha}$. Now, because of the space charge of the carriers, $\bar{E}$ must increase when the current density J increases. The results of McKay show that α increases very much more rapidly than E over a certain range; hence we may be sure that, in this range, $\bar{E}/\bar{\alpha}$, and hence $-V_a$, will *decrease* as *J increases*. The possibility of observing a negative resistance of this type renders the nature of avalanche injection worthy of the more detailed investigation described in this paper.

It may be remarked that V_a will continue to decrease with increasing current density only so long as $\bar{E}/\alpha$ decreases; if E/α goes through a minimum we may expect $\bar{E}/\bar{\alpha}$ to behave similarly. That E/α for silicon has a minimum value of about 6·3 volts is shown by the graph of figure 2, derived from the results of McKay. The available results for germanium do not extend to high enough fields for such a minimum to be observed, but there is little reason to doubt that such a minimum will occur. It therefore seems likely that the minimum value of V_a may be a constant for a given semiconductor analogous to the cathode fall of potential in a gaseous discharge, a phenomenon with which, indeed, the injecting avalanche exhibits many striking analogies, both of a formal and of a physical nature.

§ 2. EXPERIMENTAL

2.1. *Avalanche Injection Diodes with Radial Flow*

The elementary structure shown in figure 1 (*a*) does not form a practical basis for the investigation of avalanche injection because in extrinsic material the power necessary to sustain the saturated drift velocity is very much larger than can be dissipated without destructive heating effects. This difficulty may in principle be overcome by applying the electric field in brief pulses of low repetition frequency, but the pulse length which can be tolerated is so short as to make measurements difficult, particularly in a negative resistance region. If it is desired to retain the use of extrinsic material, this difficulty can be to some extent overcome by reducing the size of the non-injecting electrode so that the current flow has approximately radial symmetry in its neighbourhood and the region of saturated drift velocity is confined, even before the avalanche begins, to a thin hemispherical shell surrounding the contact. This region, in which nearly all the power is dissipated, is small enough to keep the power down to a reasonable value, and the geometry ensures that the region, being partly surrounded by an effective heat sink (the rest of the semiconductor), is as well cooled as possible.

Structures of this type, having the general form shown in figure 3 (*a*), have been made by an alloying technique, using a small particle of tin–antimony alloy to form the small n$^+$ electrode (the '*dot*'). The junction between the n$^+$ region and the n-type body of the germanium has the necessary property of excluding holes, while its size is readily made small enough to prevent excessive heating. A disadvantage of the structure is that contact to the dot can be made only with a micromanipulator.

The current–voltage characteristic measured between dot and base of such a structure has typically the form shown in figure 4. In the reverse direction (dot negative) the current at first increases rapidly with increasing voltage, but when the saturated drift velocity is reached the slope resistance becomes much greater, and continues so until very large voltages are reached. The absence of normal carrier injection ensures that the forward current at first behaves similarly†, but at about 27 volts (the *onset voltage*) there occurs a sudden decrease in slope resistance. There follows a region of negative resistance, indicated by the broken line, in which oscillations prevent accurate measurement, and finally a region in which the voltage is almost independent of current and equal to about 13 volts (the *sustaining voltage*).

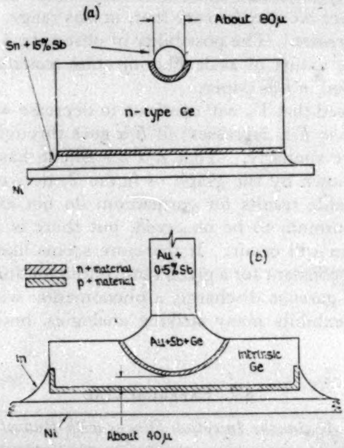


Figure 3. Practical experimental arrangements for the observation of avalanche injection.

The general shape of the forward characteristic is in keeping with the theory of avalanche injection outlined in the introduction. If this mechanism is in fact the cause, it would be expected that there would be a rapid change in the composition of the current carried by the dot as the peak of the curve is passed. That this is indeed so is indicated by the results of an ingenious experiment performed by Dr. A. F. Gibson and kindly communicated by him. In this, a pair of 10 microsecond current pulses, separated by 50 microseconds, was passed through a dot on 30 ohm cm germanium. The first, 'emitter', pulse was polarized so that the dot was biased positive with respect to the base, and its amplitude was the independent variable in the experiment. The current in the second, 'collector', pulse was maintained constant at 6 mA, and drove the dot negative with respect to the base. Because of the accumulation occurring, the voltage attained by the dot during this pulse was very sensitive to the presence of injected holes (Arthur, Gibson and Gunn 1956). Figure 5 (*b*) shows the voltage during the 'emitter' pulse and figure 5 (*a*) the voltage during the 'collector' pulse, both plotted as a function of 'emitter' current. The 'emitter' voltage curve is equivalent to the forward characteristic of figure 4, and exhibits the same sudden changes in slope. The shape of the 'collector' voltage curve shows conclusively that no injection takes place during the 'emitter' pulse until a current corresponding to the 'emitter' onset voltage is reached, and that for greater currents a large amount of injection takes place. This is exactly the result to be expected if avalanche injection is the mechanism involved, and is difficult to explain on any other hypothesis.

---

† The essential symmetry of the characteristic at low voltages is obscured in the example shown by the strong accumulation, in the reverse direction, of thermally excited carriers (Arthur, Gibson and Gunn 1956). With lower resistivity material the expected symmetry is observed.

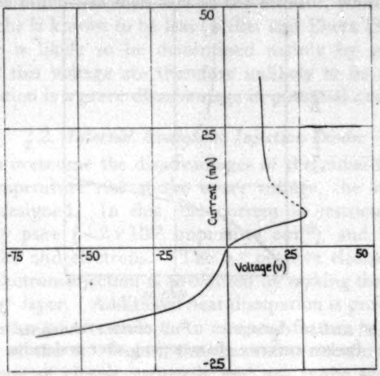
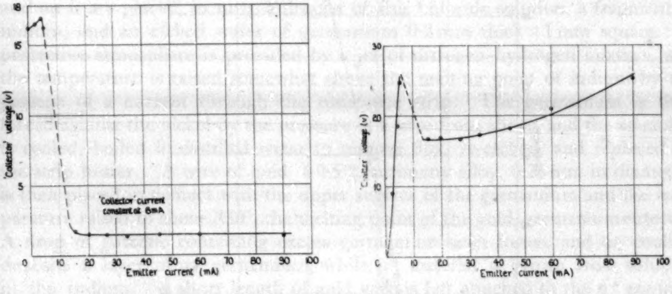


Figure 4. Typical current–voltage characteristic for a diode with radial flow.



(*a*) (*b*)

Figure 5. Demonstration of injection by an avalanche.

It is noticeable that in making oscillographic measurements on avalanche injection diodes of the alloy dot type it is frequently possible for the voltage attained on increasing the current for the first time to exceed significantly the onset voltage measured on subsequent occasions if avalanche injection has occurred meanwhile. A further decrease in onset voltage may occur if the current excursion is increased into the constant voltage region. In the limit the negative resistance region may disappear altogether and the sustaining voltage may reach abnormally low values. This effect, which is illustrated in figure 6, can be reversed, and the original characteristic approximately restored, by re-etching in hydrogen peroxide. It is therefore concluded that the changes responsible take place in the surface of the germanium. Since measurements made with

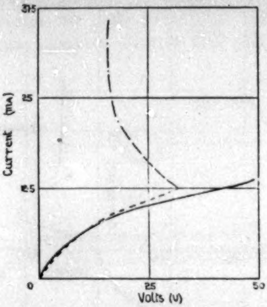


Figure 6. Example of surface damage. Full curve : characteristic before avalanche. Broken curves : characteristic after avalanche.

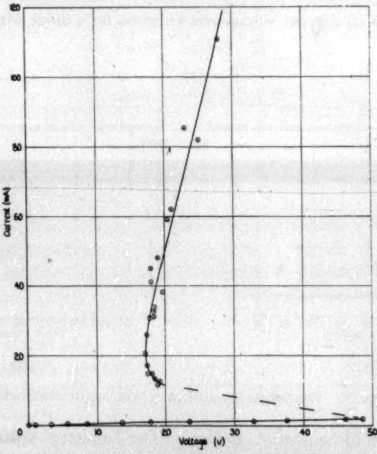


Figure 7. Pulse characteristic of a radial-flow diode.

short pulses do not show this effect, it is likely that the changes are the result of intense local heating. Figure 7 shows the result of pulse measurements (also due to Dr. Gibson) on a typical dot. Apart from the inclusion of an abnormally high passive series resistance resulting from a slightly unusual geometry, it probably represents quite well the isothermal behaviour of this type of diode.

Because the electric field at the semiconductor surface can, in a radial-flow structure, never be much less than that in the interior, while the field required to start an avalanche is known to be less (Miller and Ebers 1955), it follows that the onset voltage is likely to be determined mainly by surface conditions. Measurements of this voltage are therefore unlikely to be significant, and its uncontrolled variation is a grave disadvantage in potential circuit applications.

2.2. *Internal Avalanche Injection Diodes*

In an effort to overcome the disadvantages of the radial-flow structure, and to reduce the temperature rise at the onset voltage, the structure shown in figure 3 (*b*) was designed. In this, the current is restricted by making the germanium highly pure ($\sim 2 \times 10^{12}$ impurities cm^{-3}), and by preventing the entry of both holes and electrons. The n^+ positive electrode excludes holes as before, while electron injection is prevented by making the negative electrode in the form of a p^+ layer. Additional heat dissipation is provided by the nickel sheet attached to the p^+ electrode. An essential feature is that, as a result of the convex shape of the n^+ region, the maximum electric field occurs in the interior, at the point of closest approach, and not at the surface. It is found convenient to construct such an 'internal' diode as follows:

A cleaned piece of sheet nickel is laid on a strip of 'Kanthal' resistance alloy, and on it are placed, in turn, a droplet of zinc chloride solution, a fragment of indium, and an etched wafer of germanium 0·2 mm thick × 1 mm square. A protective atmosphere is provided by a jet of nitrogen–hydrogen mixture, and the temperature is raised somewhat above the melting point of indium by the passage of a current through the resistance strip. The germanium is then forced against the nickel by the pressure of a wire from above, and the assembly is cooled, boiled in distilled water to remove flux, re-etched, and replaced on the strip heater. A wire of gold +0·5% antimony alloy, 0·25 mm in diameter, is then placed in contact with the upper surface of the germanium and the temperature raised to above 330°, the melting point of the gold–germanium eutectic. A drop of eutectic containing excess germanium soon forms, and on cooling deposits a layer of n^+ germanium, while p^+ material is grown from solution in the indium. A short length of gold wire is left attached to the n^+ contact, so that the resulting diode has, in addition to the desired internal structure, a convenient mechanical form.

It may be remarked that the expected mode of operation of this diode is rather different from that of the alloy dot type. When the n^+ electrode is biased positively with respect to the p^+ by a small amount, the situation is just analogous to a p–i–n diode biased in reverse, and the small current which should flow results from the finite leakage velocity of the electrodes (Arthur, Gibson and Gunn 1956) and from the thermal excitation of carriers within the intrinsic region. At some larger voltage, the field between the electrodes at their point of closest approach will become large enough for avalanche injection to take place, and a negative resistance region should be observed, followed by a region of constant voltage.

It should be noted that with this arrangement two avalanches, one at each electrode, are necessary to inject both holes and electrons into the intrinsic region. The sustaining voltage for the internal diode is therefore expected to be about twice that for the radial type.

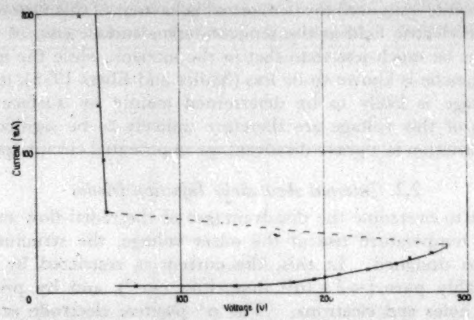


Figure 8. Typical pulse characteristic of an internal avalanche injection diode.

That these theoretical expectations are borne out in practice is shown by the pulsed current–voltage characteristic of figure 8. The increased onset voltage and the decreased current at low voltages, compared with figure 4, are at once apparent. Measurements with currents between 50 and 200 mA were difficult because of the frequent occurrence of relaxation oscillations. The voltage remained at about 30 v for currents between 200 mA and 320 mA. This value of the sustaining voltage is in satisfactory ratio with that of 13–17 v observed for the radial-flow structure.

2.3. *Measurements of Switching Time*

An avalanche injection diode which is connected to a variable voltage source of relatively low impedance will remain stable until the onset voltage is attained. At this point, however, the voltage will fall suddenly to the sustaining value, the current rising meanwhile to a large value determined mainly by the internal impedance of the source. It is of interest, in connection with possible circuit applications, to know how rapidly these changes take place and how large a current will be passed. In order to settle these points a pulse generator, using an avalanche injection diode, was constructed as shown in figure 9. In this a primary pulse about 0·3 microsecond long derived from a hydrogen thyratron is used to charge a short length of 75 Ω coaxial cable through a relatively high resistance. Little current flows through the avalanche injection diode connected effectively in parallel with the cable until the onset voltage is reached; the current then rises to a large value limited by the total effective series resistance of 150 Ω until the cable is discharged, when the voltage falls to zero. The voltage across the 75 Ω load resistor is observed by connecting across it the Y plates of a high speed cathode-ray tube capable of resolving times less than 10^{-9} second. It is readily shown that, if the length of the pulse forming cable is increased from zero, the amplitude of the pulse first becomes independent of the length of cable and its top becomes flat when the delay time from end to end of the cable equals the time of rise of the current. This was found to occur, for a typical internal diode, with a length of cable between 0·5 m and 0·75 m; the rise time, calculated from the known propagation velocity of $\frac{2}{3}c$, is thus about 3×10^{-9} sec. The current during the pulse was about 1 A. It must be remarked, however, that continued operation under these conditions frequently leads to a deterioration in the diode's properties.

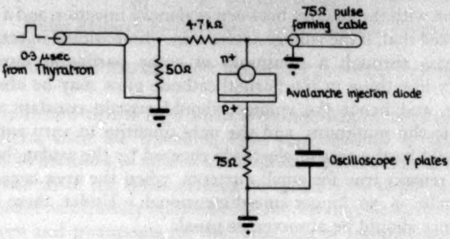


Figure 9. Circuit for the measurement of rise time.

§ 3. DISCUSSION

The experiments described above provide strong support for the belief that avalanche injection has been observed, and that it is a mechanism which may be of importance in the explanation of certain electrical phenomena of the solid state. The possibility that the observed current–voltage curves might be the result of some thermal phenomenon is disproved by the extreme rapidity of the effect; it is difficult, also, to account for the approximately constant sustaining voltage on this hypothesis. Among the implications of this result is the important deduction that the passage of an appropriately directed current density of sufficiently large magnitude through a contact to a semiconductor must result in the injection of minority carriers, *irrespective of the nature of the contact*, provided only that the drift velocity saturates at high electric fields. Under these conditions, fields high enough to initiate avalanche injection will arise unless minority carriers are injected by some other mechanism.

Indeed, it can be shown as a corollary that any sufficiently small contact must act as a rectifier, irrespective of its nature, for it has been shown theoretically by the author (1952), and experimentally by Paige (private communication), that the resistance of such a contact in the non-injecting direction is always high once the saturated drift velocity is attained. The mechanism of such a rectifier may therefore be quite independent of the presence of p–n junctions or barriers of the conventional kind. We may suspect that avalanche injection plays some part in the mechanism of such phenomena as the emission of recombination radiation in silicon carbide (Lehovec *et al.* 1951), some types of point-contact rectification and electroluminescence in phosphor powders. In all of these a large current density is associated with carrier injection and with a voltage drop which is not too small.

It is natural to enquire what is the relation between the avalanche injection diodes described above and the so-called 'positive-gap diode' of Reeves and Cooke (1955). Since this diode is manufactured by electrically forming a point contact between a silver–antimony wire and a surface of n-type germanium, it appears at first sight that avalanche injection might be the mechanism here also. It is difficult, however, to account for the very low 'sustaining' voltage observed in these diodes, though it may be that surface avalanche effects play some part.

In connection with the analogy between avalanche injection and a gas discharge it may be remarked that, if the voltage across a one-dimensional injecting avalanche does in fact pass through a minimum at some particular current density, phenomena very like those in the normal cathode glow may be observed. The current density, and hence the voltage, should remain constant at the values corresponding to the minimum, and the only quantity to vary with changes in the current should be the area of electrode covered by the avalanche. This will not, of course, remain true for small currents, when the area becomes so small that the avalanche is no longer one-dimensional. Under these conditions a negative resistance should be observed as usual.

The avalanche injection diode is potentially useful in a number of circuit applications. Apart from its use as a switching element, in which role its speed of response is at least an order of magnitude greater than that of any other semiconductor device, it is likely that use could be made of its negative resistance characteristic in oscillators and in the dissected amplifier of Shockley and Mason (1954).

In addition, if means can be found to trigger the avalanche by means of a third electrode, a number of functions at present performed by thyratrons and similar gas discharge tubes might be performed by devices employing avalanche injection. Finally, it is possible that the low slope resistance encountered at large avalanche currents, together with the reproducible voltage drop found in this region, might be of use in voltage stabilizing circuits.

ACKNOWLEDGMENTS

The author is indebted to his colleagues at the Radar Research Establishment for experimental assistance, and to Dr. A. F. Gibson for communicating figures 5 and 7 and for many helpful discussions.

This paper is published by permission of the Controller, H.M. Stationery Office.

REFERENCES

ARTHUR, J. B., GIBSON, A. F., and GUNN, J. B., 1956, *Proc. Phys. Soc.* B, **69**, 697.
GUNN, J. B., 1952, *Proc. Phys. Soc.* B, **65**, 908; 1956. *J. Electronics*, in the press.
LEHOVEC, K., ACCARDO, C. A., and JAMGOCHIAN, E., 1951, *Phys. Rev.*, **83**, 603.
MCKAY, K. G., 1954, *Phys. Rev.*, **94**, 877.
MCKAY, K. G., and MCAFEE, K. B., 1953, *Phys. Rev.*, **91**, 1079.
MILLER, S. L., 1955, *Phys. Rev.*, **95**, 1234.
MILLER, S. L., and EBERS, J. J., 1955, *Bell Syst. Tech. J.*, **34**, 883.
REEVES, A. H., and COOKE, R. B. W., 1955, *El. Commun.* **32**, 112.
RYDER, E. J., 1953, *Phys. Rev.*, **90**, 766.
SHOCKLEY, W., 1951, *Bell Syst. Tech. J.*, **30**, 990.
SHOCKLEY, W., and MASON, W. P., 1954, *J. Appl. Phys.*, **25**, 677.
WOLFF, P. A., 1954, *Phys. Rev.*, **95**, 1415.
YAMASHITA, J., and WATANABE, M., 1954, *Progr. Theor. Phys.*, **12**, 443.

On the Measurement of Minority Carrier Lifetimes in Silicon

By C. A. HOGARTH
Radar Research Establishment, Great Malvern, Worcs.

MS. received 23*rd December* 1955, *and in final form* 22*nd February* 1956

Abstract. It is shown that if high surface recombination rates can be corrected for or eliminated, then the travelling light spot method is very suitable for measurements of lifetimes in silicon since no ambiguities due to trapping effects are involved and since the measurement can readily be made on large sections of ingots instead of on carefully fashioned rods of rectangular section. Experimental results are given and treatments for the reduction of recombination velocities at surfaces of both n- and p-type silicon are described.

---

§ 1. INTRODUCTION

THE determination of τ the lifetime of minority carriers injected into silicon has been shown by Arthur, Bardsley, Gibson and Hogarth (1955) to present many difficulties which do not arise when comparable room temperature experiments are carried out with germanium crystals. One major difficulty with measurements on silicon is that in many crystals trapping effects dominate the recombination processes so that the characteristic time determined from this process is not the true lifetime. In principle it is possible to eliminate the effects of trapping by continuous illumination of the crystal with radiation in the absorption band so as to maintain the traps in a filled condition. In practice very high light intensities are often required and the resultant signals, in whatever experimental method of observing a recombination process is used, may be very small indeed. In addition, the high levels of absorbed radiation cause difficulties with temperature stabilization, and furthermore the resistance of the specimen may be markedly changed by the background illumination.

Some of the difficulties which occur when lifetime values for n-type silicon are measured by various methods have been discussed by Arthur, Bardsley, Gibson and Hogarth (1955).

In the present paper alternative methods for the elimination of trapping and high surface recombination rates are discussed.

§ 2. EXPERIMENTAL METHODS

Three methods will be discussed, two of them briefly, and the travelling light spot experiment in more detail.

(i) *The drift experiment*. Provided that the crystal filament is well etched and that the dimensions of the small sides are at least two or three diffusion lengths, this experiment should yield the true lifetime independent of surface effects. Any trapping effects can be observed as an asymmetry of the collector wave form and eliminated by continuous illumination.

(ii) *The decay of photoconductivity*. The type of experiment in which the 'on' time of the light pulse is sufficiently long for the traps to become completely filled was shown by Arthur *et al.* (1955) to be unsuitable for the determination of carrier lifetime in silicon since the long period trapping effects mask the true decay rate. If, however, by means of a spark-gap source, a very short period of excitation (of order 10^{-6} sec) is used the initial decay will yield a fairly reliable value of τ (see for example Hornbeck and Haynes 1955). During the excitation period there is little time for trapping to take place so that only a very small fraction of the total carriers become trapped and although trapping can take place during the decay period, the escape of carriers from traps does not dominate the decay. It was pointed out by Haynes (private communication) that a ready check on the validity of the result is obtained if the crystal is warmed by air from a blower. If the initial decay remains constant or gives a larger value of τ this result suggests that the effects of trapping are small.

The method requires the use of well-etched crystals of uniform cross section fitted with non-rectifying end connections. If the end connections rectify even slightly, anomalous decay curves can be obtained. In some cases the photocurrent has shown an increase with time after an initial decay. Otherwise the method is a simple one to use, but there may be a slight ambiguity in the results.

(iii) *The travelling light spot experiment.* A method for which trapping processes have no effect on the measured distribution of charge carriers is desirable. Such a method is the travelling light spot experiment introduced by Goucher (1951) and for which details have been given by Valdes (1952). This method has the advantage that it can be used on massive portions of crystals of irregular shape with one surface ground flat. Although reliable for the determination of τ in germanium there are several difficulties to be overcome when measurements are made on silicon.

(*a*) Collector non-linearity. By continuous illumination of the collector it can be forced to operate in the range where its conductance is a linear function of local minority carrier density.

(*b*) Back-scattered light. If thin specimens are used there is a danger that incident light of wavelengths at or beyond the absorption edge may be scattered from the back surface, return to the front surface of the crystal, and generate charge carriers there. This effect would give an anomalously long value of diffusion length, *but is only important in crystals for which the dimensions of the short sides are about* 1 *mm or less*, so that the filter recommended by Arthur *et al.* is not normally necessary.

(*c*) Surface recombination. With high quality silicon crystals of the type now available the large surface recombination velocity s causes a considerable modification to the distribution of injected carriers from the incident light spot. This effect becomes appreciable in germanium for lifetime values of about 100 μsec and above and the usual linear relation between the logarithm of the photovoltage and the linear distance is replaced by a curve in which even the tangent at the lower ends yields a value for diffusion length, and hence lifetime, which is lower than the true one. For silicon the values of s are much higher (typically 1000 to 5000 cm sec^{-1}) and the deviations from a linear relation are apparent for true lifetimes of about 1–2 μsec or greater.

Van Roosbroeck (1955) has treated the problem theoretically for the case of a large plane surface and has shown how corrections may in principle be applied. Cronemeyer (1955, private communication) has computed families of curves to be expected for log photovoltage against distance for true values of the diffusion length $L_{n,p}$ each family being computed for a given value of s. If s and μ the drift mobility are known then an experimental curve can be compared with an appropriate theoretical curve and $L_{n,p}$ thereby determined.

In general it is preferable to reduce the value of s to negligible proportions or to render the high value of s ineffective so that a linear relationship between log photovoltage and the distance is obtained. Means of achieving this condition for p-type silicon were suggested by a process developed by the Radio Corporation of America (Nelson 1955) for reducing the reverse currents in their fused junction antimony-alloy silicon diodes. A 2% solution of sodium dichromate in water was painted on to the crystal surface and allowed to dry. By reducing the surface recombination rate the reverse current through the diode was markedly reduced. In this laboratory it was noted that sodium dichromate is a strong oxidizing agent and it would therefore be expected to capture electrons from the surface region of the p-type silicon with which it reacted and make the surface layer more heavily p-type (known as p^+). From studies such as those of Moore and Webster (1955) it is known that a p–p^+ transition near a crystal surface is expected to lead to a reduced value of s for the surface. Thus the rapid diffusion to the surface caused by the usual rapid surface recombination would be suppressed and the charge carrier distribution would be much less disturbed.

By analogy with the treatment for p-type silicon, the surface of n-type silicon crystals should be treated with a strong reducing agent to produce an n–n^+ transition near the surface. Successful results have been obtained with 2% aqueous solutions of stannous chloride, sodium sulphite, and potassium ferrocyanide.

§ 3. EXPERIMENTAL RESULTS

Typical results will be presented for the travelling light spot experiment showing how the values derived by correcting the results obtained for crystals etched with a suitable acid etch† show good agreement with those obtained from the same crystal surfaces after treatment with oxidizing or reducing agents as described above. In some cases the results were checked by the use of the short period excitation photoconductivity decay apparatus and by the drift experiment, and consistent results obtained.

Many crystals with values of τ varying from 5×10^{-7} to 3×10^{-4} second have been examined. The experiments were carried out using white light chopped at 800 c/s and a tuned amplifier. The specimens were thick enough to eliminate the effects of back-scattered light, had one surface ground flat and were either treated with an acid etch and the results corrected, or treated with the appropriate oxidizing agent (p-type) or reducing agent (n-type) and the results analysed in the usual manner. A continuous background illumination was used in each case. Values for the diffusion constant were taken as $D_n = 35\ \mathrm{cm^2\,sec^{-1}}$ for electrons and $D_p = 9\ \mathrm{cm^2\,sec^{-1}}$ for holes. From measurements in this laboratory and elsewhere it was found that the value of s for silicon treated with the acid etch referred to above was always of order $2500\ \mathrm{cm\,sec^{-1}}$. Since the correction to the log photovoltage against distance curves was not very sensitive to small changes in s, this mean value was used in the application of the correction curves. The consistency of the results with those from other methods used on the same crystal justified this procedure.

---

† The most suitable acid etch for silicon seems to be a mixture of 4 parts HNO_3 with 1 part HF.

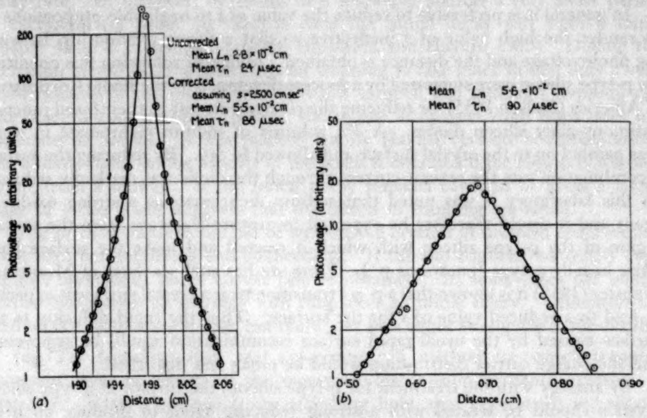


Figure 1. (*a*) Measurement on p-type silicon crystal etched with acid; (*b*) measurement on p-type silicon crystal treated with sodium dichromate.

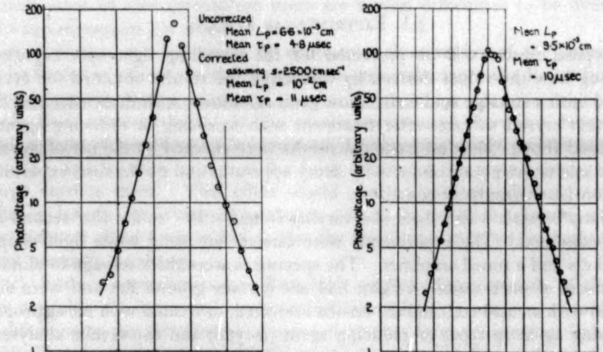


Figure 2. (*a*) Measurement on n-type silicon crystal etched with acid; (*b*) measurement on n-type silicon crystal treated with sodium sulphite.

A typical result for a p-type single crystal after acid etching is shown in figure 1 (*a*). Correcting for surface recombination, values of $L_n = 0{\cdot}055$ cm and $\tau_n = 86 \times 10^{-6}$ sec are obtained. Figure 1 (*b*) shows the results obtained on the same crystal after treatment of the surface with sodium dichromate. In this case the curves are like those traditionally obtained for germanium under conditions where surface recombination can be neglected. The mean value of L_n is 0·056 cm and hence $\tau_n = 90 \times 10^{-6}$ sec. For this particular crystal the lifetime was also determined by the drift mobility type of experiment and a value of 80×10^{-6} sec was obtained for the crystal in filamentary form with cross section about 2·5 mm × 2·5 mm but with no special surface treatment other than a normal acid etch. Measurement of the initial photoconductive decay also gave a value of 80×10^{-6} sec for the lifetime. It is interesting to note that had the tail ends of the curves in figure 1 (*a*) been assumed to be linear and surface recombination negligible in the traditional manner, a value of only 0·028 cm would be obtained for L_n corresponding to a value of $\tau_n = 24 \times 10^{-6}$ second.

Comparable results are obtained for n-type material. Figure 2 (*a*) shows the distribution curve obtained for a crystal treated wiht an acid etch, and figure 2 (*b*) the curve obtained after treatment of the crystal surface with sodium sulphite solution. Excellent agreement is obtained between the value of τ computed from figure 2 (*a*) using the correction curves, namely 11 microseconds, and the value calculated directly from figure 2 (*b*), 10 microseconds. The value obtained from the initial decay of photoconductivity using the spark-gap apparatus was also 11 microseconds.

§ 4. Conclusions

Of the methods available for the accurate determination of minority carrier lifetime in silicon, the drift experiment and the study of photoconductive decay using a spark-discharge source are reliable, practically unambiguous with regard to errors arising from trapping effects and, once the measuring equipment is assembled, simple to carry out. The methods do however necessitate the cutting of rods of uniform cross section, soldering on two base leads, and very careful etching of the whole crystal. They are not suitable for the examination of large ingots.

The travelling light spot provides a ready method for the examination of regions of large ingots necessitating only grinding flat the area of surface to be investigated, fitting one base lead, and etching. Precautions have to be taken to correct for or to eliminate the effects of surface recombination and to linearize the collector. If these precautions are taken, the method provides a simple and accurate measurement of diffusion length and hence lifetime, in which the effects of carrier trapping are not observed.

Acknowledgments

It is a pleasure to acknowledge the assistance received from colleagues in this laboratory and from Dr. D. C. Cronemeyer for early communication of his surface recombination calculations, and also the gift of high quality silicon crystals from the Philco Research Laboratories and the Bell Telephone Laboratories. Acknowledgment is made to the Controller, H.M. Stationery Office, for permission to publish this paper.

References

Arthur, J. B., Bardsley, W., Gibson, A. F., and Hogarth, C. A., 1955, *Proc. Phys. Soc.* B, **68,** 121.

Goucher, F. S., 1951, *Phys. Rev.*, **81,** 475.

Hornbeck, H., and Haynes, J. R., 1955, *Phys. Rev.*, **97,** 311.

Moore, A. R., and Webster, W. M., 1955, *Proc. Inst. Radio Engrs, N.Y.*, **43,** 427.

Nelson, H., 1955, communication to the I.R.E./A.I.E.E. Semiconductor Device Research Conference.

van Roosbroeck, W., 1955, *J. Appl. Phys.*, **26,** 380.

Valdes, L. B., 1952, *Proc. Inst. Radio Engrs, N.Y.*, **40,** 445.

Electrical Measurements on Natural Galena at Low Temperatures

By D. M. FINLAYSON and D. GREIG
Department of Natural Philosophy, University of Aberdeen

MS. received 13*th February* 1956

Abstract. The Hall effect and resistivity of n-type single crystals of natural galena containing from 6×10^{15} to 10^{19} extrinsic electrons per cm^3 have been measured from room temperature down to liquid helium temperatures. No appreciable change in the number of extrinsic electrons has been observed down to 4°K in any crystal.

While the mobility is proportional to $T^{-5/2}$ down to a temperature of about 70°K, the low temperature mobility varies with the number of extrinsic electrons in a manner which cannot be explained by current theories of scattering.

It is estimated that at room temperature a purely intrinsic crystal of PbS would contain 2×10^{15} conduction electrons per cm^3.

---

§ 1. INTRODUCTION

EXTENSIVE measurements of the electrical properties of lead sulphide have been made at temperatures above that of liquid air (see, for example, Putley 1952, 1955, Petritz and Scanlon 1955). In the low temperature region isolated measurements have been made by Putley at 20°K, and resistivity measurements on very impure material have been reported by Dunaev (1947). Meanwhile, recent theoretical work (e.g. Howarth and Sondheimer 1953) suggests that at low temperatures, polar crystals may differ markedly in their behaviour from the better known elemental semiconductors. Following on infra-red work carried out in this laboratory (Paul and Jones 1953), a supply of natural galena covering a considerable impurity range was available. It seemed, therefore, worth while to extend electrical measurements to low temperatures.

§ 2. EXPERIMENTAL METHODS

2.1. *Low Temperature Apparatus*

A miniature hydrogen–helium liquefier of the type constructed by Chester and Jones (1953) was used to produce low temperatures. Samples were enclosed in a small copper can which served as the bulb of a gas thermometer. This can was surrounded by an outer can in good thermal contact with the helium reservoir. The intervening space could be evacuated or filled with a small quantity of helium gas, so that the sample could make thermal contact with, or be isolated from, the helium bath at will. A thermostat-controlled heating coil, wound on the inner can, allowed temperatures between the boiling points of helium, hydrogen and oxygen to be obtained.

2.2. *Selection and Mounting of Samples*

Specimens were obtained by selection from a number of sources of natural galena, by measuring the Hall effect and resistivity at room temperature. The measurements were all made on single crystals. Samples containing 10^{17} to 10^{19} electrons per cm^3 could readily be found, and after examination of a large number of specimens, a few were obtained containing about 10^{16} electrons per cm^3, the purest having 6×10^{15} electrons per cm^3. (In the infra-red the latter showed the smallest absorption coefficient we have measured so far but, apart from this, no detailed correspondence between infra-red and electrical measurements was observed.)

Rectangular samples were cleaved from larger blocks to give dimensions about 1 cm × 0·3 cm × 0·1 cm and were checked for homogeneity by Hall and resistance measurements using pressure probes. For low temperature measurements six 41 s.w.g. copper wires were welded on to the sample, using the condenser discharge technique described by Mitchell (1954). The current leads were welded to the ends, and two leads welded to each side were arranged symmetrically, so as to give two resistance and two Hall readings. Good low resistance welds could be obtained by discharging a 75 μF condenser charged to about 30 v. On cooling, the contact resistances often tended to increase considerably. This could be remedied during a run by discharging a condenser across the offending, and any one of the other leads, the condenser being charged to about 10 v. This process did not give rise to any observable change in the crystal's characteristics.

2.3. *Electrical Measurements*

The resistivity and Hall effect were measured by the normal d.c. methods, using a Tinsley No. 3184 potentiometer. All measurements were taken for both directions of the current, while the Hall readings were also taken for both directions of the field. A Tickford Type A magnet was used, giving a maximum field of 5500 oersted, for a pole gap of 5 cm.

§ 3. EXPERIMENTAL RESULTS

3.1. *Hall Effect*

The Hall constant is given by $R = 10^8 \times Et/HIf$ cm^3 coulomb^{-1}, where E is the measured e.m.f., t the thickness, H the field, I the current, and f is a potential distribution correction factor, discussed by Volger (1950) and Frank (1953). H was measured with an accuracy to 2%, while E and I were measured to closer than 1%, giving a relative error in R of less than 5%. However, values of E obtained from different parts of the crystal sometimes varied appreciably, due presumably to inhomogeneity, thus setting a limit of within 10 to 15% to the absolute accuracy of those measurements. The Hall coefficient is plotted against temperature in figure 1.

3.2. *Resistivity*

The resistivity is given by $\rho = twE/dI$, where w is the width and d is the distance between the probes. The relative accuracy of the measurements was to about 2%. Values calculated from the opposite sides of the crystal sometimes differed appreciably, so that the absolute accuracy was again probably within 10 to 15%. Resistivity is plotted against temperature in figure 2.

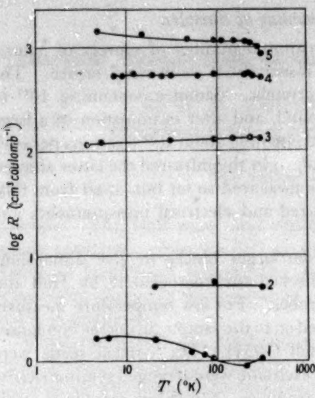


Figure 1.

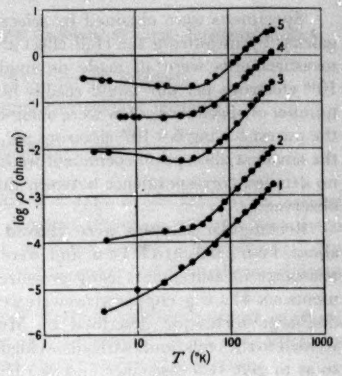


Figure 2.

3.3. *Mobility*

The mobility is obtained from the resistivity and Hall coefficients by the relation $\mu = R/r\rho$, where r is a factor which varies between 1 and 1·9 (see, for

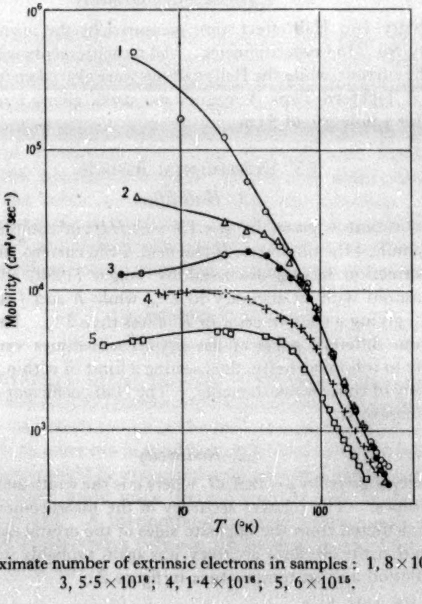


Figure 3. Approximate number of extrinsic electrons in samples: 1, 8×10^{18}; 2, $1{\cdot}3\times10^{18}$; 3, $5{\cdot}5\times10^{16}$; 4, $1{\cdot}4\times10^{16}$; 5, 6×10^{15}.

example, Johnson and Lark-Horovitz 1950, Jones 1951), depending on the electron density and the scattering mechanism. In polar semiconductors, both optical and acoustical scattering may be important, but the evaluation of r for such mixed scattering is complex. Considering acoustical scattering only, r varies between 1·00 and 1·17, going from complete degeneracy to the classical limit. We have estimated that the value of r which would occur, if the scattering were due to optical modes alone, would differ from the acoustical values we have used by 5% at the classical limit, and would become identical at complete degeneracy. Uncertainty in the true value of r is likely to introduce, therefore, a very small error in the mobility. Mobility is plotted against temperature in figure 3.

§ 4. DISCUSSION

4.1. *Hall Effect*

It will be seen from figure 1 that sample 5 has a sufficiently small concentration of extrinsic electrons to show intrinsic behaviour at room temperature, the Hall coefficient increasing down to about 200°K. Taking the room temperature value of the Hall constant, and the extrinsic electron density, and assuming a mobility ratio of 1·4 (Brebrick and Scanlon 1954), we estimate that material which is intrinsic at room temperature should contain about 2×10^{15} electrons per cm^3.

Below about 200°K the Hall coefficient is practically constant with temperature for all samples. Since the dielectric constant of lead sulphide is 15·3, we might expect an application of the 'hydrogen model' to the calculation of the impurity activation energy to yield a value of the same order of magnitude as germanium (0·01 ev). Below 50°K sample 5 shows a small increase in the Hall constant apparently corresponding to a small decrease in the number of conduction electrons. Evaluation of the Fermi level indicates that classical approximations can be applied. The calculation of an activation energy from the number of conduction electrons is ambiguous unless the acceptor concentration is known, but every simple model seems to yield a negative activation energy. We conclude that in lead sulphide containing 6×10^{15} electrons per cm^3 no separation between impurity levels and conduction band can be observed.

4.2. *Mobility*

In the region between room and liquid air temperatures, all samples have a mobility obeying the relation $\mu = AT^{-5/2}$ reported by Petritz and Scanlon (1955) and by Putley (1952). Smith (1954) reports that if the mobility law at intermediate temperatures is of the above form then, in the values of A, there is scatter which bears no relation to the number of extrinsic impurities. This is confirmed by Putley (1955), who quotes values of the mobility ranging from 140 to 870 cm^2 v^{-1} sec^{-1}.† Petritz and Scanlon claim that there is no noticeable dependence of mobility on impurity concentration, and that the mobility in the intermediate temperature range is an intrinsic property of lead sulphide. We find that in the temperature range down to 70°K the mobility of the purest sample (5) is about half that of samples 1, 2 and 3, while sample 4 occupies an intermediate position. A closer inspection of the curves presented by Petritz and

---

† We understand from a private communication that this scatter is greatly reduced if natural single crystals are used.

Scanlon indicates a similar change in mobility between their most and least pure samples. With our range of mobilities at room temperature, the intrinsic resistivity of lead sulphide should lie between 3·8 and 7 ohm cm. The room temperature resistivity of our purest sample was 4·3 ohm cm.

At lower temperatures, for all samples, the mobility deviates considerably from the $T^{-5/2}$ law, this deviation beginning at about 70°K in every case. By analogy with elemental semiconductors, this might be expected as a result of scattering by ionized impurities. First attempts to apply the Conwell–Weisskopf (1950) relation, and its modification for degenerate samples due to Johnson and Lark-Horovitz (1947), gave no numerical agreement. As a wider range of impurity concentrations was used, it became apparent that an increase in conduction electrons was accompanied by an increase in the mobility, and a sample containing 10^{19} electrons per cm^3 deviated least from the $T^{-5/2}$ law. It is difficult to see how any combination of lattice and impurity scattering, neutral or ionized, could give this effect.

To explain a maximum in the Hall effect of germanium, Hung (1950) suggested that impurity atoms could, instead of existing as discrete centres, form a narrow band below the conduction band. The width of this band should increase as the number of impurities increases, and so the mobility in the impurity band might be expected to increase with impurity concentration. A similar picture might be applied to lead sulphide, if we suppose that at low temperatures the conduction electrons occupy such a band. Then in this temperature region the mobility will be determined mainly by the impurity concentration, while ordinary lattice and Conwell–Weisskopf scattering can be ignored, since they will only be important for electrons in the main conduction band. No adequate theoretical explanation for the existence of impurity band conduction has been given, and so no attempt at estimating the temperature dependence of impurity mobility can be made. It may be noted, however, that the purest samples in figure 3 show a decrease in mobility at the lowest temperatures.

The Hung model may also supply an explanation of the failure of the 'hydrogen model' to predict a value of impurity activation energy. At the lowest impurity concentration of 6×10^{15} electrons per cm^3 the impurity band may be of sufficient width to overlap the conduction band, and no activation energy will be observed.

Scattering by optical modes of vibration in polar crystals has been discussed at length by Howarth and Sondheimer (1953). An attempt has been made by Petritz and Scanlon (1955) to combine this optical mode scattering with the well-known acoustical scattering for the case of lead sulphide below its Debye temperature (about 200°K). Their experimental results, extending down to 77°K, gave only a narrow temperature range where agreement was possible. Our results indicate that below about 70°K the mobility is controlled by the impurity concentration, and so lattice scattering, both acoustical and optical, is completely masked. Consequently, lead sulphide appears to be an unsuitable substance for the verification of the Howarth–Sondheimer theory as it is developed at present. At higher temperatures the $T^{-5/2}$ law for the mobility is well established experimentally, but no satisfactory theory has yet been presented.

ACKNOWLEDGMENTS

We are very grateful to Professor R. V. Jones who provided the research facilities and obtained the specimens of galena. We would like to thank the Caird Trust and the Department of Scientific and Industrial Research for financial assistance. Mr. R. Wildgoose assisted with the construction of the liquefier.

REFERENCES

BREBRICK, R. F., and SCANLON, W. W., 1954, *Phys. Rev.*, **96**, 598.
CHESTER, P. F., and JONES, G. O., 1953, *Proc. Phys. Soc.* B, **66**, 296.
CONWELL, E., and WEISSKOPF, V. F., 1950, *Phys. Rev.*, **77**, 388.
DUNAEV, J. A., 1947, *C.R. Acad. Sci. U.R.S.S.*, **55**, 21.
FRANK, V., 1953, *Appl. Sci. Res.* B, **3**, 129.
HOWARTH, D. J., and SONDHEIMER, E. H., 1953, *Proc. Roy. Soc.* A, **219**, 53.
HUNG, C. S., 1950, *Phys. Rev.*, **79**, 727.
JOHNSON, V. A., and LARK-HOROVITZ, K., 1947, *Phys. Rev.*, **71**, 374; 1950, *Ibid.*, **79**, 176.
JONES, H., 1951, *Phys. Rev.*, **81**, 149.
MITCHELL, W. H., 1954, *J. Sci. Instrum.*, **31**, 147.
PAUL, W., and JONES, R. V., 1953, *Proc. Phys. Soc.* B, **66**, 194.
PETRITZ, R. L., and SCANLON, W. W., 1955, *Phys. Rev.*, **97**, 1620.
PUTLEY, E. H., 1952, *Proc. Phys. Soc.* B, **65**, 388; 1955, *Ibid.*, **68**, 22.
SMITH, R. A., 1954, *Physica*, **20**, 910.
VOLGER, J., 1950, *Phys. Rev.*, **79**, 1023.

The Electrical Conductivity of Composite Media

By E. H. KERNER

Physics Department, University of Buffalo, New York

MS. received 30*th September* 1955 *and in revised form* 20*th February* 1956

Abstract. A general theory is given of the gross electrical conductivity (or thermal conductivity, dielectric constant or magnetic permeability) of a composite medium having its isotropic granular components distributed spatially in a disordered way. It is shown that the gross conductivity is a weighted superposition of the component conductivities with weights containing the volume fractions of the components and an intensity factor depending on the average electric field for the whole medium and for each component. By introducing a suitably defined 'average electrostatic problem', the case of grains spherical in the mean and suspended in a uniform medium is examined in detail, and the gross conductivity explicitly found. The limiting case in which the suspending fluid is withdrawn and the grains coalesce is shown to give rise to a discontinuity in the gross conductivity. Coated grains, e.g. oxide-coated metallic grains, are also treated.

§ 1. Introduction

In this and a succeeding paper (Kerner 1956) a solution is proposed to the very old problem of how to deduce some macroscopic physical properties of media made up, on a relatively microscopic scale, of several pure materials having specified properties and aggregated in various disordered ways. Such media include suspensions of powders in some uniform substrate, suitably bonded aggregates of powders, and the like. The conductive properties, comprising dielectric constant, magnetic permeability, and electrical and thermal conductivities, may be treated together because of the identical form of the conduction laws controlling them. Here we shall use the language of electrical conductivity in discussing these properties; with simple changes in nomenclature the results may be applied to the other cases. The elastic properties of composites are the subject of the companion paper (Kerner 1956).

In a general way it will be shown that the gross conductivity of a large mass of composite is a weighted linear superposition of the component conductivities. The weights are the product of a geometrical factor—the fraction of the total volume occupied by a component species—and an 'intensity' factor: the ratio of (the average value in a component species of the component of the electric field in the direction of the externally applied field) to (the average value in the whole medium of the same quantity). The geometrical factor is assumed to be known. The evaluation of the intensity factor requires a special statistical hypothesis, of a type introduced by Bruggeman (1935), which leads in the case of grains suspended in some medium, to the picture of an average grain surrounded by a shell of the suspending medium. Sufficiently far beyond this shell there lies on the average a uniform medium with a conductivity equal to the desired gross conductivity; but between this shell and the uniform average medium there must be, for elementary reasons of continuity, some intermediate zone of variable conductivity. Fortunately, this zone can be by-passed and the gross conductivity found by a simple application of electrostatic theory. The case of packed grains becomes a special case of suspended grains by allowing the suspending fluid to disappear.

The recognition of this intermediate zone appears to be essential for a completely consistent description of the gross conductivity, and is the element in the present theory which is absent in previous investigations.

One of the earliest discussions of the bulk properties of a composite medium seems to be due to Rayleigh (1892) who investigated the conductivity of a medium containing a second substance as a cubic array of spheres. Bruggeman (1935) in a fundamental paper studied the dielectric constant of various kinds of binary composites and reviewed the older literature. Our results for packed aggregates of grains generalize his. For suspensions of grains in a uniform medium we propose an analysis different from Bruggeman's, one leading in a natural way to the Lorentz–Lorenz law, with which Bruggeman's results are in conflict. In addition, Bruggeman's method applied to a suspension of different species of grains fails to give the correct result in the limit that the suspending fluid vanishes and the grains become packed. A discussion by Polder and Van Santen (1946) of grain suspensions appears to be based on Bruggeman's statistical hypothesis for packed grains. More recently Landauer (1952) has verified, in a comparison with experimental data, Bruggeman's result for binary packed powder aggregates.

§ 2. GENERAL CONSIDERATIONS

The composite medium will be taken to be made up of grains of materials labelled 2, 3,... suspended in a fluid labelled 1. The label 0 is used for the composite as a whole. Each component will be assumed to be an isotropic, homogeneous substance. The composite medium itself will be assumed to be homogeneous on a gross scale, i.e. on a scale of dimensions large compared with grain dimensions and grain separations. We make also the basic assumption that the grains have no spatial ordering, but are intermixed spatially in a haphazard manner.

Consider an infinite slab of composite medium of width very large compared with grain size, and bounded on either face by wide slabs of some uniform medium having the conductivity σ. The extreme end faces of this sandwich are held at some fixed potential difference and a steady current flows through it. This situation is what is often realized in the measurement of the conductivity σ_0. Choosing a z axis perpendicular to the faces of the sandwich, for two equal long segments L_0, L of it deep in the interiors of σ_0 and one slab of σ and far from all interfaces of the sandwich, we must have

$$\sigma_0/\sigma = \int_L E_z\,dz \Big/ \int_{L_0} E_{0z}\,dz,$$

or

$$\sigma E_z = \sigma_0\left(\bar{E}_{1z}\frac{L_1}{L_0} + \bar{E}_{2z}\frac{L_2}{L_0} + \ldots\right), \qquad \ldots\ldots(1)$$

where L_1/L_0, $L_2/L_0\ldots$, are the fractions of L_0 that are intercepted in media 1, 2,..., and lie in these media. Here use has been made of the mean value theorem in the form

$$\int_{L_0} E_{0z}\,dz = \int_{L_1} E_{1z}\,dz + \int_{L_2} E_{2z}\,dz + \ldots = \bar{E}_{1z}L_1 + \bar{E}_{2z}L_2 + \ldots \qquad \ldots\ldots(2)$$

The field E_z is the constant one in σ at points far from the interface of σ and σ_0.

Now using Gauss' theorem for a cylinder Z parallel to the z axis and having ends S_0, S deep in σ_0 and one slab of σ we get

$$\int_Z j_n\,dS = 0 = \int_S \sigma E_z\,dS - \int_{S_0} j_z\,dS_0,$$

the contribution to the surface integral from the walls parallel to z being zero on account of the homogeneity of the composite. Hence, breaking the last integral up into integrals over S_1, $S_2, \dots$, the total areas cut out on S_0 by media $1, 2, \dots$, we have

$$\sigma E_z = \sigma_1 \bar{E}_{1z} S_1/S_0 + \sigma_2 \bar{E}_{2z} S_2/S_0 + \dots \qquad \dots\dots(3)$$

Now S_i/S_0 and L_i/L_0 are equal and, in fact, the same as the fraction of some large volume τ_0 occupied by material i for the reason that a large plane surface, a long line segment, and a large volume all effectively sample the medium in the same way. Calling these common ratios v_i it follows from equations (1)–(3) that

$$\sigma_0 = \sum \sigma_i v_i \bar{E}_{iz}/\bar{E}_{0z}. \qquad \dots\dots(4)$$

We have also $\Sigma v_i = 1$ and by equations (2),

$$\sum v_i \bar{E}_{iz}/\bar{E}_{0z} = 1, \qquad \dots\dots(5)$$

so that the coefficients of σ_i in equation (4) are of the nature of true statistical weights. These are the basic equations for what follows.

Though the components of a composite medium are isotropic, homogeneous substances, the medium clearly will not be isotropic in the large if the grains are, for example, ellipsoidal and preferentially oriented in some fixed direction. If $\sigma_0{}^*$ be the conductivity dyadic of the composite medium, the preceding analysis holds intact when σ_0 is replaced by $\hat{z} \cdot \sigma_0{}^* \cdot \hat{z}$. Then $\hat{z}$ is to be chosen in three successive steps to be along the principal axes of $\sigma_0{}^*$ so as to obtain in three coupled equations the three principal conductivities of the medium. These axes are determined by the statistics and geometries of the grains of various species.

To compute the field-ratio factors, let us go from grain to grain of a given species i in order to make an average of the grain sizes and shapes and of the fields they contain. Also around each grain let some surface S, of a shape roughly similar to that of the grain but of larger dimensions, be constructed. If S is sufficiently small, say S_1, then the material between the grain surface and S_1 will be some layer of the suspending fluid 1. If S is sufficiently large, say S_2, the material beyond S_2 will in the mean be a medium having the conductivity σ_0 of the composite, and the field beyond S_2 will average to the value $\bar{E}_{0z}\hat{z}$. This suggests, as a general model for finding $\bar{E}_{iz}/\bar{E}_{0z}$, the 'average electrostatic problem' of a grain σ_i bounded by a layer σ_1, then by some variable-conductivity layer, having a conductivity ranging from σ_1 to σ_0, and finally by a uniform medium containing at infinity a uniform field.

Though they are difficult to define exactly, it is clear that some maximal S_1 and minimal S_2 exist. That S_1 and S_2 cannot coincide may be shown from the example of spherical grains—a direct computation of $\bar{E}_{iz}/\bar{E}_{0z}$ assuming σ jumps from σ_1 to σ_0 at a single boundary allows σ_0 to be found independently from equations (4) and (5); the two values disagree.

To avoid the apparently complex intermediate zone between S_1 and S_2 we need simply eliminate $\bar{E}_{0z}$ from equations (4) and (5),

$$\sigma_0 = \sum \sigma_i v_i \bar{E}_{iz}/\bar{E}_{1z} \Big/ \sum v_i \bar{E}_{iz}/\bar{E}_{1z}, \qquad \dots\dots(6)$$

where for convenience the dimensionless ratio $\bar{E}_{iz}/\bar{E}_{1z}$ rather than $\bar{E}_{iz}$ is introduced.

§ 3. Spherical Average Grains

The simplest kind of grain consistent with gross isotropy is that which in the mean is spherical. This is also perhaps the most important case arising in practice. An elementary treatment of the 'average electrostatic problem' here yields $\bar{E}_{iz}/\bar{E}_{1z} = 3\sigma_1/(2\sigma_1+\sigma_i)$. This is independent of the grain size, the size of the layer σ_1, the size and conductivity of the intermediate zone, and the conductivity σ_0 of the exterior uniform medium. It is a consequence solely of the spherical symmetry. We see in fact that in this case we need not have assumed the model of the 'average electrostatic problem' described above, but could have employed a less restrictive assumption.

The gross conductivity is just

$$\sigma_0 = \sum \sigma_i v_i \frac{3\sigma_1}{\sigma_i+2\sigma_1} \Big/ \sum v_i \frac{3\sigma_1}{\sigma_i+2\sigma_1}, \qquad \ldots\ldots(7)$$

where the summations are over all components, including 1.

For a binary composite, grains of σ_2 suspended in the fluid σ_1, a rearrangement of equation (7) gives

$$\frac{\sigma_2/\sigma_1+2}{\sigma_2/\sigma_1-1} = v_2 \frac{\sigma_0/\sigma_1+2}{\sigma_0/\sigma_1-1}.$$

This is the Lorentz–Lorenz law $v_2(n^2+2)/(n^2-1) = \text{constant}$, when dielectric constants and refractive indices are introduced instead of conductivities, and it holds without restriction on v_2, in contrast to the result Rayleigh (1892) found for a regular cubic array of spherical grains.

§ 4. Packed Grains

Let us withdraw the suspending fluid ($v_1 \to 0$), allowing the grains to coalesce into a tight compact. According to equation (7)

$$\sigma_0 = \sum_{i=2} \sigma_i v_i \frac{3\sigma_1}{\sigma_i+2\sigma_1} \Big/ \sum_{i=2} v_i \frac{3\sigma_1}{\sigma_i+2\sigma_1} \equiv \Sigma_2(\sigma_1)/\Sigma_1(\sigma_1), \qquad \ldots\ldots(8)$$

which states that σ_0 still depends on σ_1, a material now absent from the composite.

This seeming paradox has the following answer. Equation (7) may be written, with the terms in v_1 isolated from the rest, as $v_1(\sigma_0-\sigma_1) = \Sigma_2 - \sigma_0\Sigma_1$. This equation is satisfied identically if we *choose* σ_1 to be equal to σ_0, and determine σ_0 according to $\sigma_0 = \Sigma_2(\sigma_0)/\Sigma_1(\sigma_0)$; and it is satisfied *independently of the value of* v_1. In particular it holds for $v_1 = 0$. Physically what this means is that if we take the compacted composite of conductivity σ_0 and prepare a uniform fluid of this same conductivity, then we may inject the fluid into the compact, causing the grains to become separated but leaving the gross conductivity unchanged at the value σ_0 no matter how much fluid is injected. Only the fluid of conductivity determined by

$$\sum_{i=1} v_i \frac{\sigma_0-\sigma_i}{2\sigma_0+\sigma_i} = 0 = \sigma_0\Sigma_1(\sigma_0) - \Sigma_2(\sigma_0) \qquad \ldots\ldots(9)$$

has this property.

The σ_0's defined by equations (8) and (9) are not the same, and should not be the same, because at $v_1 = 0$, σ_0 must in general be discontinuous. Consider, for example, perfectly conducting grains suspended in a perfectly insulating fluid. For any finite amount of fluid, no matter how small, granted that the grains are suspended one must find $\sigma_0 = 0$, in accordance with equation (8). But for a strictly zero amount of fluid the conductivity must jump suddenly to infinity, in accordance with equation (9). In other words, equation (8) correctly provides σ_0, exclusive only of the physically meaningful finite change in conductivity accompanied by an infinitesimal change in v_1 at $v_1=0$. This jump in conductivity of course is not in reality abrupt, but is continuous due to the part played by non-ohmic conduction processes, which have been neglected.

The final result, equation (9), for a compact of grains provides a polynomial of degree n in σ_0 for an n-component compact. The right-hand side of the equation is a monotonic function of σ_0, increasing uniformly from the value -1 to $+\frac{1}{2}$ as σ_0 ranges from zero to infinity. Hence there is, as there must be, a unique positive σ_0 satisfying it. The equation includes as a special case Bruggeman's (1935) and Landauer's (1952) result for a binary composite, which has been checked against experiment.

It is interesting to observe that equation (9) follows from each of equations (4) and (8) separately if one assumes that the average grain in the compact is bounded directly by the uniform average medium σ_0. In this case $\overline{E}_{iz}/\overline{E}_{0z}=3\sigma_0/(2\sigma_0+\sigma_i)$, independently of the grain size. This means that in the compact two 'grains' which happen to be adjacent can be considered a single 'grain'. The 'grains' in other words have no well-defined identity in the compact. It is only through the procedure of coalescing a suspension of grains that the 'graininess' of the compact has a meaning.

§ 5. COATED GRAINS

In important cases, for example in composites of oxide-coated metallic grains, the composite conductivity is strongly influenced by even thin layers of foreign material on the grain surfaces. To the extent that such layers may be described as having a definite conductivity, we can compute the gross conductivity by a slight extension of the earlier discussion. We need only keep in mind that the summations in equation (6) include the grain cores as well as the grain coatings.

For average spherical grains σ_i, with average spherical coatings σ_i', immersed in the fluid σ_1, it is straightforward to show that

$$\begin{aligned} \overline{E}_{iz}&=L_{1i}\overline{E}_{1z}, & \overline{E}_{iz}'&=L_{1i}'\overline{E}_{1z} \\ L_{1i}&\equiv 9\sigma_1\sigma_i'/D_{1i}, & L_{1i}'&=3\sigma_1(\sigma_i+2\sigma_i')/D_{1i} \\ D_{1i}&\equiv(\sigma_i+2\sigma_i')(\sigma_i'+2\sigma_1)+2h_i(\sigma_1-\sigma_i')(\sigma_i'-\sigma_i) \end{aligned}$$

where h_i denotes the ratio of the volume of the core to the total volume of core plus coating. Equation (6) gives for the overall conductivity

$$\sigma_0=\sum(\sigma_i v_i L_{1i}+\sigma_i' v_i' L_{1i}')\Big/\sum(v_i L_{1i}+v_i' L_{1i}')$$

with the conventions $L_{11}=1$, $L_{11}'=0$. It is easily checked that when the cores or coatings are shrunk to zero, or are made identical, or that when the coatings are given the conductivity σ_1, equation (6) is reproduced.

Forcing the coated grains into coalescence by taking $v_1=0$ gives the same paradox, which has the same answer, as earlier. The result, analogous to equation (9), is that the compacted conductivity is governed by

$$0=\sum_{i=2}[9v_i\sigma_i'(\sigma_i'-\sigma_0)+3v_i'(2\sigma_i'+\sigma_i)(\sigma_i'-\sigma_0)]/D_{0i}, \qquad \ldots\ldots(10)$$

where D_{0i} stands for D_{1i} with σ_1 replaced by σ_0. This equation is readily shown to determine a unique positive σ_0. Having coalesced the coated grains we may make all the grain coatings identical. The resulting suspension of cores in this grain-coating fluid should have its conductivity described by equation (7). A somewhat involved algebraic manipulation does indeed show that σ_0 according to equation (7) identically satisfies equation (10) when all σ_i' in the latter have a common value.

The extension of the foregoing discussion to two-dimensional composites, consisting of long parallel cylinders of the various components, is readily performed and need not be detailed here.

ACKNOWLEDGMENT

The author has profited from several valuable discussions with Dr. S. Mrozowski. The work was partially supported by the U.S. Atomic Energy Commission.

REFERENCES

BRUGGEMAN, D. A. G., 1935, *Ann. Phys., Lpz.*, **24**, 635.
KERNER, E. H., 1956, *Proc. Phys. Soc.* B, **69**, 808.
LANDAUER, R., 1952, *J. Appl. Phys.*, **23**, 779.
POLDER, D., and VAN SANTEN, J. H., 1946, *Physica*, **12**, 257.
RAYLEIGH, LORD, 1892, *Phil. Mag.*, **34**, 481.

The Elastic and Thermo-elastic Properties of Composite Media

By E. H. KERNER
Department of Physics, University of Buffalo, New York

MS. received 30*th September* 1955 *and in revised form* 20*th February* 1956

Abstract. In a continuation of a previous paper, it is shown here how the gross bulk and shear moduli of a composite material consisting of a suspension of grains or a compact of grains may be deduced. The grains are assumed to be perfectly bonded to the suspending medium or to each other, and are taken to be spheres in the mean. By using an averaging procedure due to Bruggeman, and analysing the effect of a uniform hydrostatic compression and of a uniform tension on an average grain, a pair of de-coupled equations for the gross moduli is found for suspensions. When the suspending medium vanishes and the grains are packed, these equations become coupled and there is exhibited a discontinuity in the gross moduli. The bulk coefficients of linear expansion of the two kinds of composites are found from an analysis of the dilatation and bulk stress for average spherical grains when the composite as a whole is subjected to some small temperature change. All results are free of any limitation on the number of components.

§ 1. Introduction

In an accompanying paper (Kerner 1956) is described a method for deducing gross conductive properties—like electrical conductivity—of certain kinds of composites, namely, of compacts of grains and suspensions of grains. The present report aims to cover the elastic and thermo-elastic properties.

The two elastic constants, shear modulus and bulk modulus, of a macroscopically isotropic and homogeneous composite will be found in terms of the moduli and concentrations of its components. These components will be assumed to be in the form of grains suspended in and bonded to some uniform suspending medium. It is assumed also that the grains are distributed spatially at random, and that in the mean they are spherical. The latter does not imply for many applications too stringent a restriction on the sizes and shapes of individual grains. It appears to be the simplest assumption compatible with gross isotropy. The case of tightly packed grains is then treated by allowing the suspending fluid to vanish. Finally a computation of the gross thermal expansion coefficient of a composite is undertaken along lines similar to those developed for the elastic moduli.

As was mentioned in the analysis of the conductive properties, we may construct an average grain surrounded by an average shell of suspending medium; sufficiently far beyond this we have the average medium. But in between there is some intermediate zone having properties shading continuously, in some unknown way, from those of the suspending medium to those of the average medium. Fortunately, in very much the same way as in the conduction problem this intermediate zone need not be analysed. By averaging both the dilatation and bulk stress in a uniform hydrostatic compression, the intermediate zone may be by-passed, so to speak; and similarly, in a simple tension, by averaging both the elongation and the normal stress across a plane orthogonal to the applied tension, the properties of the intermediate zone are eliminated. The result is a pair of de-coupled equations for the desired gross moduli. In taking the limit that the grains become packed these equations become coupled.

Though there are many data on the elastic properties of various composites, it appears that hardly any of them are useful for comparing with theoretical results —the data are very scattered, and most often too rough and incomplete. A systematic experimental study to test the theoretical conclusions would be highly desirable.

The most important previous paper in this field seems to be that of Bruggeman (1937) where references to earlier literature may be found. Bruggeman, among other things, analysed binary compacts and suspensions using, as he did also for flow properties (Bruggeman 1935), an averaging procedure substantially equivalent to the one we employ. One of our pair of equations describing compacts of grains includes and generalizes one of Bruggeman's. For the rest, although our approach to the problem is within the spirit of the averaging procedure advanced by Bruggeman, the analysis and results are quite different. In both the conduction and elasticity problems, Bruggeman's results for suspensions fail to include as a special case the results for packed grains. In addition, the method used by Bruggeman for both elastic and conduction properties of suspensions fails—in the analysis of conduction properties—to give the correct result, the Lorentz–Lorenz law; it is doubtful, because of this, that the corresponding analysis of elastic properties is valid other than as an approximation.

More recently, Mackenzie (1950), using a 'self-consistent' method due to Fröhlich and Sack (1946), has analysed the elastic properties of a material having a suspension of air bubbles in it. This method ignores the intermediate zone mentioned above, and is valid only in the limit of small total pore volume in spite of the fact that it gives the correct result in the limit that the whole volume is pore volume.

§ 2. SUSPENDED GRAINS

We label quantities referring to any grain species i with the subscript i, reserving the subscript 0 for the composite as a whole; the index 1 refers to the suspending fluid.

Let a large mass of the composite be subjected to a simple hydrostatic compression. The mass behaves elastically in the large like some uniform material having a shear modulus μ_0 and bulk modulus k_0 which we wish to find. Though the state of stress and strain at any point within the composite be exceedingly difficult to find, we assume that as we go from grain to grain of any species i there is some average state of stress and strain in and near the average i-grain that has a simple character: one calculable by examining the average spherical i-grain, characterized by μ_i, k_i, bounded by a layer μ_1, k_1, then the intermediate zone, and finally the average medium μ_0, k_0. We construct, in short, an 'average elastic problem' similar to the average problem described in the conductivity paper.

In any case we must have that the total change in volume in the composite must be the sum of the volume changes in its different parts. Denoting the dilatation by l this means

$$\int l_0\,d\tau_0 = \sum \int l_i\,d\tau_i \qquad \ldots\ldots(1)$$

or

$$\bar{l}_0 = \sum \bar{l}_i v_i \qquad \ldots\ldots(2)$$

where $v_i = \tau_i/\tau_0 =$ fraction of total volume occupied by material of species i, and $\Sigma v_i = 1$.

Let us denote by θ the bulk stress, giving the mean pressure at a point. For any plane surface drawn in the composite we must have that the mean normal force across it is the sum of the mean normal forces across its parts which lie in the different grains and the suspending fluid. That is, analogously to equation (1) we must have

$$\int \theta_0\,dS_0 = \sum \int \theta_i\,dS_i \qquad \ldots\ldots(3)$$

or, like equation (2),

$$\bar{\theta}_0 = \sum \bar{\theta}_i v_i \qquad (v_i = S_i/S_0 = \tau_i/\tau_0). \qquad \ldots\ldots(4)$$

Now, in the average elastic problem, for a simple compression the radial displacements are of the form

$$u_r = A_i r \text{ (in } i\text{)}; \quad u_r = A_1 r + B_1 r^{-2} \text{ (in 1)}; \quad u_r = A_0 r + B_0 r^{-2} \text{ (in 0)}.$$

The dilatation is $3A_i = \bar{l}_i$ in each medium, and the bulk stress is $3k_iA_i = \bar{\theta}_i$ in each. When these are introduced into equations (2) and (4) we may eliminate A_0 between them and express k_0 in terms of A_i, or better, A_i/A_1. The latter ratio is fixed by the continuity conditions on displacements and tractions at the interface of the average i-grain sphere and the layer of suspending fluid 1, and has the value $(3k_1+4\mu_1)/(3k_i+4\mu_1)$. The calculation of k_0 then gives

$$k_0 = \sum \frac{k_i v_i}{3k_i+4\mu_1} \Big/ \sum \frac{v_i}{3k_i+4\mu_1}. \qquad \ldots\ldots(5)$$

To find the gross shear modulus μ_0 we analyse the average effect of a uniform tension applied to the composite. The average elastic problem then consists of a spherical inclusion μ_i, k_i, with its 1-layer and intermediate zone, embedded in μ_0, k_0 in which there is a uniform tension, say in the z direction, at infinity. A problem of this sort, one having the inclusion embedded directly in a second medium, has been discussed by Goodier (1933).

The use of a dilatation equation, like equation (1), or a bulk stress equation, like equation (3), will provide no new information here. For, by the superposition principle any uniform state of stress may be derived from a system of uniform tensions applied in suitable directions; in particular a uniform (negative) hydrostatic compression can be obtained by superposing uniform tensions in three mutually orthogonal directions, and a calculation of dilatations for equation (1) from any one of these tensions is the same as a calculation for all three simultaneously, as may be verified directly. We could have, indeed, dispensed with a separate derivation of equation (5) via the uniform hydrostatic compression, and relied only on a dilatation calculation from the uniform tension. The superposition principle tells us, in short, that the result in equation (5) is independent of what uniform stress we apply to the composite; invoking the hydrostatic compression was a matter of convenience, not necessity.

Consider a long line segment L_0 parallel to the z axis in the composite under tension. The elongation l_{zz} measures the change in length per unit length in the z direction, and bears a relationship to the state of tension similar to that which the dilatation bears to the state of hydrostatic compression. Analogously to equation (1) we may write

$$\int l_{zz0}\,dz_0 = \sum \int l_{zzi}\,dz_i,$$

and by the mean value theorem again,

$$\bar{l}_{zz0} = \sum \bar{l}_{zzi} v_i, \qquad \ldots\ldots(6)$$

(v_i is L_i/L_0 as well as S_i/S_0 and τ_i/τ_0, L_i/L_0 denoting the fraction of L_0 lying within material i).

If T_{zz} denotes the normal stress across an x, y plane in the composite, we may, in a parallel with equation (3), equate the normal force across this plane to the sum of the forces across its parts lying in the different components:

$$\int T_{zz0}\,dS_0 = \sum \int T_{zzi}\,dS_i$$

and thence

$$\bar{T}_{zz0} = \sum \bar{T}_{zzi} v_i. \qquad \ldots\ldots(7)$$

If we extend Goodier's analysis now to the average elastic problem, the displacements are, in spherical coordinates (σ denoting Poisson's ratio and T_0 the uniform tension)

$$\text{in } i \quad u_r = F_i r + D_i r + 2\sigma_i E_i r^3 + (3D_i r + 6\sigma_i E_i r^3)\cos 2\theta,$$

$$u_\theta = -(3D_i r + (7-4\sigma_i)E_i r^3)\sin 2\theta.$$

$$\text{in } 1 \quad u_r = -\frac{A_1}{r^2} - \frac{3B_1}{r^4} + \left(\frac{5-4\sigma_1}{1-2\sigma_1}\frac{C_1}{r^2} - \frac{9B_1}{r^4}\right)\cos 2\theta + F_1 r + D_1 r + 2\sigma_1 E_1 r^3 + (3D_1 r + 6\sigma_1 E_1 r^3)\cos 2\theta,$$

$$u_\theta = -\left(\frac{2C_1}{r^2} + \frac{6B_1}{r^4}\right)\sin 2\theta - \{3D_1 r + (7-4\sigma_1)E_1 r^3\}\sin 2\theta.$$

$$\text{in } 0 \quad u_r = -\frac{A_0}{r^2} - \frac{3B_0}{r^4} + \left(\frac{5-4\sigma_0}{1-2\sigma_0}\frac{C_0}{r^2} - \frac{9B_0}{r^4}\right)\cos 2\theta + \frac{T_0}{4\mu_0} r \left(\frac{1-\sigma_0}{1+\sigma_0} + \cos 2\theta\right),$$

$$u_\theta = -\left(\frac{2C_0}{r^2} + \frac{6B_0}{r^4}\right)\sin 2\theta - \frac{T_0}{4\mu_0} r \sin 2\theta.$$

The direct calculation of l_{zz} and of $T_{zz} = \lambda l + 2\mu l_{zz}$ (λ being the Lamé coefficient $2\sigma\mu/(1-2\sigma)$) gives, when substituted into equations (6) and (7),

$$F_0 + 4D_0 = \sum \left(F_i + 4D_i + \frac{28}{5} E_i a_i^2\right) v_i,$$

$$(3\lambda_0 + 2\mu_0)F_0 + 8\mu_0 D_0 = \sum \left[(3\lambda_i + 2\mu_i)F_i + 8\mu_i\left(D_i + \frac{7}{5}E_i a_i^2\right)\right] v_i,$$

where a_i is the radius of the average i-grain. However, $3F$ is the dilatation and $3F(3\lambda + 2\mu)$ is the bulk stress, and so, according to our discussion above of the superposition principle, we have already, in equations (1) and (3), effectively placed $F_0 = \Sigma F_i v_i$ and $(3\lambda_0 + 2\mu_0)F_0 = \Sigma(3\lambda_i + 2\mu_i)F_i$. Hence

$$D_0 = \sum \left(D_i + \frac{7}{5}E_i a_i^2\right) v_i, \qquad \mu_0 D_0 = \sum \mu_i \left(D_i + \frac{7}{5}E_i a_i^2\right) v_i,$$

or, dividing these by $D_1+\frac{7}{5}E_1a_i^2$ (which is independent of a_i) and eliminating between them the ratio $(D_1+\frac{7}{5}E_1a_i^2)/D_0$, we get

$$\mu_0=\frac{\Sigma\mu_iR_iv_i}{\Sigma R_iv_i},\qquad R_i=\frac{D_i+\frac{7}{5}E_ia_i^2}{D_1+\frac{7}{5}E_1a_i^2}.$$

The usual boundary conditions at $r=a_i$ give, after a long calculation,

$$R_i=\frac{15\mu_1(1-\sigma_1)}{(7-5\sigma_1)\mu_1+(8-10\sigma_1)\mu_i},$$

whence

$$\mu_0=\mu_1\frac{\displaystyle\sum{}'\frac{\mu_iv_i}{(7-5\sigma_1)\mu_1+(8-10\sigma_1)\mu_i}+\frac{v_1}{15(1-\sigma_1)}}{\displaystyle\sum{}'\frac{\mu_1v_i}{(7-5\sigma_1)\mu_1+(8-10\sigma_1)\mu_i}+\frac{v_1}{15(1-\sigma_1)}}. \qquad \ldots\ldots(8)$$

The Σ' stands for a summation excluding the index 1.

§ 3. PACKED GRAINS

We examine now the limiting case of the final formulae (5) and (8) for k_0 and μ_0 in the limit that the suspending fluid vanishes and the grains become packed. Placing $v_1=0$ in these formulae gives again the same 'paradox' noted in the conductivity problem, that the gross moduli depend still on the properties of the vanished fluid 1. However, exactly the same answer to the difficulty prevails in the present case. This is that, upon isolating the terms in v_1 in equations (5) and (8), these equations are identically satisfied for all v_1 (including $v_1=0$), provided that k_1, μ_1 are chosen to be k_0, μ_0 and that the latter simultaneously satisfy

$$0=\sum{}'\frac{k_i-k_0}{3k_i+4\mu_0}v_i, \qquad \ldots\ldots\ (9)$$

$$0=\sum{}'\frac{(\mu_0-\mu_i)v_i}{(7-5\sigma_0)\mu_0+(8-10\sigma_0)\mu_i}. \qquad \ldots\ldots(10)$$

In effect, in going to $v_1=0$ the fluid 1 is to be replaced by fluid 0.

The packed-grain moduli are given now by the coupled equations (9) and (10). The discontinuity in these moduli with those given by equations (5) and (8) for any arbitrarily small v_1 corresponds to a real physical discontinuity which must in fact be accounted for by a theory of composite media. Consider, for example, a composite of liquid drops suspended in some very rigid material. So long as there is a finite amount of the latter the gross shear modulus remains high, but as soon as it disappears the modulus drops to zero; this is correctly described by the above equations.

It is of interest to note that had we considered the packed grains bounded directly by the average medium k_0, μ_0 we would find equation (9) coming independently from (2) and from (4) and equation (10) independently from (6) and from (7).

§ 4. THERMO-ELASTIC PROPERTIES

In the suspended-grain composite let a small temperature drop T place the system in a stressed state as compared with its original state, taken to be unstressed. The composite over-all behaves like some uniform medium having an expansion coefficient γ_0 and suffering a dilatation γ_0T. Constructing as before the average thermo-elastic problem, we have the displacements,

$$\begin{aligned} &\text{in } i \quad u_r=\tfrac{1}{3}\gamma_i Tr+A_i r,\\ &\text{in } 1 \quad u_r=\tfrac{1}{3}\gamma_1 Tr+A_1 r+B_1 r^{-2},\\ &\text{in } 0 \quad u_r=\tfrac{1}{3}\gamma_0 Tr+B_0 r^{-2}. \end{aligned}$$

The first terms are the thermal, the remaining terms the mechanical parts of the displacements.

The continuity conditions at the average *i*-sphere surface give

$$A_i=\left[\tfrac{1}{3}(\gamma_1-\gamma_i)T+A_1\left(1+\tfrac{3}{4}\frac{k_1}{\mu_1}\right)\right]\Big/\left(1+\tfrac{3}{4}\frac{k_i}{\mu_1}\right),$$

while the dilatation and bulk stress equations (1) and (3) are

$$\gamma_0 T=\Sigma(\gamma_i T+3A_i)v_i; \qquad 0=\Sigma k_i A_i v_i.$$

From the latter equation we get, using A_i above,

$$3A_1=\sum\frac{k_i(\gamma_i-\gamma_1)Tv_i}{1+\frac{3}{4}k_i/\mu_1}\Big/\sum k_i v_i\frac{1+\frac{3}{4}k_1/\mu_1}{1+\frac{3}{4}k_i/\mu_1}.$$

This introduced into the former equation yields a term involving the combination of elastic moduli defining k_0 according to equation (5). The result is, after a reduction,

$$\gamma_0=\sum\gamma_i v_i+4\frac{\mu_1}{k_0}\sum\frac{k_0-k_i}{4\mu_1+3k_i}(\gamma_1-\gamma_i)v_i.$$

Now turning to the limit $v_1=0$, we find that γ_0 still depends apparently on the properties of the annihilated suspending fluid 1. However, as was noted before, we must take $\mu_1=\mu_0$ in this limit; and then the coefficient of γ_1 is zero by virtue of equation (9), giving for the expansion coefficient of the packed aggregate

$$\gamma_0=\sum{}'\gamma_i v_i+4\frac{\mu_0}{k_0}\sum{}'\frac{k_i-k_0}{4\mu_0+3k_i}\gamma_i v_i=\left(4\frac{\mu_0}{k_0}+3\right)\sum{}'\frac{\gamma_i v_i}{4\mu_0/k_i+3}.$$

It is seen that only in the case that all components have the same bulk modulus does the simple 'rule of mixtures' $\gamma_0=\Sigma\gamma_i v_i$, hold.

ACKNOWLEDGMENTS

I am indebted to Dr. S. Mrozowski for bringing these problems to my notice and for valuable critical remarks. The work was partially supported by the U.S. Atomic Energy Commission.

REFERENCES

BRUGGEMAN, D. A. G., 1935, *Ann. Phys. Lpz.*, **24**, 635; 1937, *Ibid.*, **29**, 160.
FRÖHLICH, H., and SACK, R., 1946, *Proc. Roy. Soc.* A, **185**, 415.
GOODIER, J. N., 1933, *Trans. Amer. Soc. Mech. Engrs*, **55**, 39.
KERNER, E. H., 1956, *Proc. Phys. Soc.* B, **69**, 802.
MACKENZIE, J. K., 1950, *Proc. Phys. Soc.* B, **63**, 2.

Ultrasonic Relaxation Theory for Liquids

By J. H. ANDREAE† AND J. LAMB‡

† Akers Research Laboratories, Imperial Chemical Industries Ltd., Welwyn, Herts.
‡ Electrical Engineering Department, Imperial College.

MS. received 2nd August 1955, *and in revised form 9th January* 1956

Abstract. The propagation of a sound wave through a liquid can be described in terms of the absorption per wavelength μ and the velocity c. It is generally accepted that an important part of the absorption is due to the relaxation of internal degrees of freedom, and the aim of this paper is to derive relationships between the acoustic quantities μ and c and the thermodynamic parameters which describe specific molecular processes.

The analysis is based on an expansion of the adiabatic compressibility. It is valid for the general case of structural relaxation involving changes both in internal energy and in volume as a molecule passes from one stable equilibrium state to another. Attention is given to the combined effects of more than one relaxing process, where each process has a different relaxation time. The equations are derived in a form suitable for the analysis of experimental results.

§ 1. INTRODUCTION

It is generally accepted that an important part of the ultrasonic absorption in liquids is due to the relaxation of internal degrees of freedom. A finite time is required for energy to be transferred from translational kinetic energy to each of the various forms in which it can be stored internally by the molecules. When the period of a sound wave is much less than this relaxation time for a particular internal state, the latter will not contribute to the propagation of the sound wave. The energy stored internally is described as thermal when it is absorbed at constant volume, and as structural when any volume change is involved. In general, the transfer of energy to internal degrees of freedom will not take place at constant volume so that it is necessary to consider structural effects.

Kneser (1938) and Herzfeld (1941) have transferred their theories for relaxation phenomena in gases to pressure-independent problems in liquids. Also Hall (1948) has put forward a theory to account for the ultrasonic absorption in water, which he attributes to an isothermal relaxation between an ordered (ice-like) and a random structure of the water molecules. The theories of Kneser and Herzfeld are appropriate to the treatment of relaxation processes which depend only on temperature, while that of Hall is concerned with pressure-dependent processes which take place isothermally. More recently, Meixner (1952) and Davies (1954) have treated the general structural case for a single relaxation mechanism.

The aim of the present paper is to derive expressions for the ultrasonic absorption in the presence of several independent relaxing processes.

§ 2. THEORY

2.1. *The Propagation of Sound at High Frequencies*

A consideration of the equations of motion for a one-dimensional sound wave propagated in a homogeneous isotropic medium having an adiabatic compressibility β, and a density ρ, shows that the velocity of propagation will be c, where

$$c^2 = 1/\beta\rho. \qquad \ldots\ldots(1)$$

If the medium absorbs energy from the sound wave, the amplitude of, say, the pressure fluctuations will be attenuated with distance according to:

$$\text{Instantaneous pressure} = P + p\exp\{-\alpha x + j\omega(t - x/c)\}, \qquad \ldots\ldots(2)$$

where x is the distance measured in the direction of propagation, t is the time, $\omega/2\pi$ is the frequency of the sound wave, P is the mean pressure, p is the amplitude of the pressure fluctuations at $x = 0$, α is the attenuation constant, or absorption coefficient, and c is the phase velocity. The absorption coefficient and phase velocity may be combined in the usual manner in a single complex velocity C, such that

$$1/C = (1/c)(1 - j\alpha c/\omega). \qquad \ldots\ldots(3)$$

Equation (1) will be valid for this complex velocity so long as the compressibility is allowed to have an imaginary part to represent the effect of the second (dilatational) viscosity. Introducing C for c in (1) we find

$$1/C^2 = (1/c^2)(1 - j\mu/\pi) = \beta_\omega\rho, \qquad \ldots\ldots(4)$$

where the absorption per wavelength $\mu = 2\pi\alpha c/\omega$ is assumed, in the following, to be much less than unity: this is a convenient assumption which is usually valid. β_ω is the effective complex adiabatic compressibility measured at the frequency $\omega/2\pi$.

Equation (4) does not take into account the effects of shear viscosity and heat conduction. The absorption coefficients for these two processes are usually sufficiently small to be combined with the α of equation (3) by simple addition.

The object of the following theory is to relate the ultrasonic parameters μ and c to the thermodynamic parameters which describe specific molecular processes.

2.2. *The Basic Thermodynamics*

The behaviour of a liquid may be likened partly to that of a solid, in which long-range order has disappeared, and partly to that of a gas, in which intermolecular forces have become predominant. Both of these approaches to the liquid state have been made with limited success. In the following it will be assumed that a simple gas-like thermodynamic treatment can be applied to ordinary non-viscous liquids, and the theory will be derived for the particular case of a small perturbation by a compressional monotonic sound wave.

The passage of a sound wave will cause small local fluctuations about equilibrium in pressure P, volume V, temperature T, and entropy S, at the sound frequency $\omega/2\pi$. It is postulated that these fluctuations will be governed by an 'equation of state'

$$X = F_\omega(Y, Z), \qquad \ldots\ldots(5)$$

where X, Y, Z are any three of the variables P, V, T, S, and $F_\omega()$ is a function depending on ω. If Y and Z be made to vary sinusoidally with time at the frequency $\omega/2\pi$, then for small departures from equilibrium it is assumed that X will respond linearly and sinusoidally, so that

$$dX = (\partial X/\partial Y)_{Z,\omega}\,dY + (\partial X/\partial Z)_{Y,\omega}\,dZ. \qquad \ldots\ldots(6)$$

The partial derivatives in equation (6) have meaning only at a single frequency $\omega/2\pi$. Thus it is possible to form the usual relations between the common thermodynamic parameters at this frequency. The more important of the latter are $\beta_\omega = -V^{-1}(\partial V/\partial P)_{S,\omega}$, the adiabatic compressibility encountered in equation (4); $\beta_{T\omega} = -V^{-1}(\partial V/\partial P)_{T,\omega}$, the isothermal compressibility; $\theta_\omega = V^{-1}(\partial V/\partial T)_{P,\omega}$, the thermal expansion coefficient; $P_\omega{}^* = T(\partial P/\partial T)_{V,\omega}$, the internal pressure; $C_{\mathrm{p}\omega} = T(\partial S/\partial T)_{P,\omega}$, the specific heat at constant pressure; $C_{\mathrm{v}\omega} = T(\partial S/\partial T)_{V,\omega}$, the specific heat at constant volume; and $\gamma_\omega = C_{\mathrm{p}\omega}/C_{\mathrm{v}\omega}$, the ratio of the specific heats: each of these parameters is a function of the frequency.

A second postulate is made to the effect that the variation with frequency of the equation of state (5) is due only to the existence of internal degrees of freedom, and that a general equation of state,

$$X = F(Y, Z, y_1, y_2, y_3, \ldots y_m, \ldots), \qquad \ldots\ldots(7)$$

can be formed with new variables $y_1, y_2, y_3, \ldots y_m, \ldots$. Each of the y-variables represents an internal degree of freedom. For small displacements from equilibrium it follows from (7) that

$$dX = (\partial X/\partial Y)_{Z,y}\,dY + (\partial X/\partial Z)_{Y,y}\,dZ + \sum_m (\partial X/\partial y_m)_{Z,Y,m}\,dy_m, \qquad \ldots\ldots(8)$$

where the subscript y means that all the y's are held constant during differentiation, and the subscript m means that all the y's *except* y_m are held constant. Combination of equations (6) and (8) leads to the formula

$$(\partial X/\partial Z)_{Y,\omega} = (\partial X/\partial Z)_{Y,y} + \sum_m (\partial X/\partial y_m)_{Y,Z,m}(\partial y_m/\partial Z)_{Y,\omega}. \qquad \ldots\ldots(9)$$

This is the basis of the following theoretical treatment and it is used for expanding each of the thermodynamic parameters β, β_T, θ, P^*, C_{p}, and C_{v} as a sum of terms. The first term $(\partial X/\partial Z)_{Y,y}$ is independent of frequency by the second postulate, since it is unaffected by any of the internal degrees of freedom. The other terms are 'relaxing terms', being the respective contributions of each of the internal states to the parameter $(\partial X/\partial Z)_{Y,\omega}$. At sufficiently high sound frequencies none of the relaxing terms will be operative because the period of the sound wave will become much shorter than the relaxation time for the transfer of energy to each of the internal states. The parameter will then take the value $(\partial X/\partial Z)_{Y,y}$, which will be called the 'residual' or 'infinite frequency' value.

2.3. *Expansion of the Compressibility*

Inspection of equation (4) shows that the frequency dependence of the phase velocity c, and the absorption per wavelength μ, is determined entirely by that of the adiabatic compressibility β_ω. The latter may be expanded by formula (9) to give

$$\begin{aligned}\beta_\omega = -V^{-1}(\partial V/\partial P)_{S,\omega} &= -V^{-1}(\partial V/\partial P)_{S,y} - \sum_m V^{-1}(\partial V/\partial y_m)_{S,P,m}(\partial y_m/\partial P)_{S,\omega} \\ &= \beta_y + \sum_m \beta_m. \qquad \ldots\ldots(10)\end{aligned}$$

The residual compressibility β_y is independent of frequency, so that attention is directed to the relaxing terms

$$\beta_m = -V^{-1}(\partial V/\partial y_m)_{S,P,m}(\partial y_m/\partial P)_{S,\omega}. \qquad \ldots\ldots(11)$$

In order to find useful expressions for the relaxing compressibility, it is necessary to introduce additional parameters. The *m*th partial volume change is defined as $\Delta V_m = (\partial V/\partial y_m)_{P,T,m}$. The *m*th partial enthalpy change is $\Delta H_m = T(\partial S/\partial y_m)_{P,T,m}$, and the *m*th partial change of internal energy is $\Delta E_m = T(\partial S/\partial y_m)_{V,T,m}$. Following from these definitions,

$$\Delta H_m = \Delta E_m + P_y{}^*\Delta V_m. \qquad \ldots\ldots(12)$$

The specific heats are now expanded by formula (9):

$$\begin{aligned} C_{p\omega} = T(\partial S/\partial T)_{P,\omega} &= T(\partial S/\partial T)_{P,y} + \sum_m T(\partial S/\partial y_m)_{P,T,m}(\partial y_m/\partial T)_{P,\omega} \\ &= C_{py} + \sum_m C_{pm}, \qquad \ldots\ldots(13) \\ C_{v\omega} = T(\partial S/\partial T)_{V,\omega} &= T(\partial S/\partial T)_{V,y} + \sum_m T(\partial S/\partial y_m)_{V,T,m}(\partial y_m/\partial T)_{V,\omega} \\ &= C_{vy} + \sum_m C_{vm}. \qquad \ldots\ldots(14) \end{aligned}$$

It follows that the *m*th relaxing components of the specific heat, C_{pm} and C_{vm}, may be written

$$C_{pm} = \Delta H_m(\partial y_m/\partial T)_{P,\omega} \quad \text{and} \quad C_{vm} = \Delta E_m(\partial y_m/\partial T)_{V,\omega}. \qquad \ldots\ldots(15)$$

The thermal expansion θ and the internal pressure P^* may be expanded in a similar manner. Then, using standard thermodynamic relations, unwanted terms are eliminated (see Appendix) and, finally, the required equations are obtained:

$$\frac{\beta_\omega}{\beta_y} = 1 + \sum_m \frac{C_{pm}}{(\gamma_y - 1)C_{p\omega}}\left[\left(\gamma_y\frac{\Delta E_m}{\Delta H_m} - 1\right)^2 + \frac{\gamma_y{}^2}{2C_{py}}\sum_n C_{pn}\left(\frac{\Delta E_n}{\Delta H_n} - \frac{\Delta E_m}{\Delta H_m}\right)^2\right] \qquad \ldots\ldots(16)$$

$$\left(C_{vm} - \frac{\Delta E_m}{\Delta H_m}C_{pm}\right)\sum_n\left(C_{vn} - \frac{\Delta E_n}{\Delta H_n}C_{pn}\right) = (C_{p\omega} - C_{v\omega})\left[C_{pm}\left(\frac{\Delta E_m}{\Delta H_m}\right)^2 - C_{vm}\right] \qquad \ldots\ldots(17)$$

On the left-hand side of equation (16) is the factor β_ω/β_y. β_y is the compressibility at a frequency ω_y so high that there can be no absorption due to the relaxation of internal degrees of freedom. If c_y is the velocity at ω_y then equation (4) demands

$$\rho\beta_y = (1/c_y)^2(1 - j0) = (1/c_y)^2.$$

Therefore,

$$\beta_\omega/\beta_y = (c_y/c)^2(1 - j\mu/\pi), \qquad \ldots\ldots(18)$$

and $(c_y/c)^2$ is the real part of β_ω/β_y, while $-\mu/\pi$ is the imaginary part of β_ω/β_y divided by the real part.

2.4. *The Frequency Dependence of the Relaxing Specific Heat*

It has been explained in § 2.2 that each y_m in equation (7) represents an internal degree of freedom within the liquid. It is now assumed that the y_m variables are governed by rate equations like

$$\tau_m{}^{PT}\frac{dy_m}{dt} = -(y_m - \bar{y}_m{}^{PT}), \qquad \ldots\ldots(19)$$

where $\bar{y}_m{}^{PT}$ at any instant is that value of y_m which would be attained if the system were then allowed to proceed to equilibrium at constant pressure and temperature; $\tau_m{}^{PT}$ is the corresponding relaxation time. Equation (19) has been fully discussed by Davies and Lamb (1956).

For sinusoidal perturbation at the frequency $\omega/2\pi$, equation (19) becomes

$$\delta y_m = \frac{\delta\bar{y}_m{}^{PT}}{1+j\omega\tau_m{}^{PT}}. \qquad \ldots\ldots(20)$$

Each y_m is now a parameter indicating the progress of a reaction occurring within the liquid. This reaction may be the elevation of a molecule to a vibrational energy state; in this case the reactants are unexcited molecules and the products of the reaction are excited molecules. On the other hand, the reaction may represent a change in the orientation of the molecules from a random to an ordered arrangement. A 'chemical' reaction, in which the molecules exchange their constituent atoms, is a further possibility.

Comparing equation (20) with (15), it may be seen that the relaxing components of the specific heat take the form

$$C_{\mathrm{p}m} = \frac{C_{\mathrm{i}m}}{1+j\omega\tau_m}, \qquad \ldots\ldots(21)$$

where

$$C_{\mathrm{i}m} = \Delta H_m(\partial\bar{y}_m{}^{PT}/\partial T)_P \qquad \ldots\ldots(22)$$

is the zero-frequency, or static, value of $C_{\mathrm{p}m}$. The superscripts PT have been omitted from τ_m in equation (21) for convenience, and their presence will be implied in the remainder of the paper.

The third stage in the derivation of the ultrasonic absorption is now complete. Equation (4) related parameters α and c, measurable by ultrasonic techniques, with a complex thermodynamic compressibility β_ω. The latter was then converted into an expression (16) involving specific heats and heats of reaction only. The complex nature of the compressibility is realized in the frequency relations (21) for the components of the specific heat.

2.5. *The Absorption per Wavelength and the Velocity Dispersion*

Substitution of the expression for $C_{\mathrm{p}m}$ from equation (21) in equation (16) yields

$$\frac{\beta_\omega}{\beta_y} = \frac{\begin{array}{r} C_{\mathrm{p}y} + \sum\limits_m \dfrac{C_{\mathrm{i}m}}{1+\omega^2\tau_m{}^2}\left\{1+K_m+\sum\limits_n \dfrac{K_{mn}C_{\mathrm{i}n}(1-\omega^2\tau_m\tau_n)}{1+\omega^2\tau_n{}^2}\right\} \\ -j\left[\sum\limits_m \dfrac{\omega\tau_m C_{\mathrm{i}m}}{1+\omega^2\tau_m{}^2}\left\{1+K_m+2\sum\limits_n \dfrac{K_{mn}C_{\mathrm{i}n}}{1+\omega^2\tau_n{}^2}\right\}\right] \end{array}}{C_{\mathrm{p}y} + \sum\limits_m \dfrac{C_{\mathrm{i}m}}{1+\omega^2\tau_m{}^2} - j\sum\limits_m \dfrac{\omega\tau_m C_{\mathrm{i}m}}{1+\omega^2\tau_m{}^2}} \qquad \ldots\ldots(23)$$

where

$$K_m = \left[\gamma_y\frac{\Delta E_m}{\Delta H_m} - 1\right]^2\Big/(\gamma_y - 1) \quad \text{and} \quad K_{mn} = \frac{\gamma_y}{2(C_{\mathrm{p}y} - C_{\mathrm{v}y})}\left[\frac{\Delta E_m}{\Delta H_m} - \frac{\Delta E_n}{\Delta H_n}\right]^2. \qquad \ldots\ldots(24)$$

Equation (23) is of the form

$$\beta_\omega/\beta_y = \frac{P-jQ}{R-jS} = \frac{PR+QS}{R^2+S^2}\left[1-j\left(\frac{QR-PS}{PR+QS}\right)\right], \qquad \ldots\ldots(25)$$

in which P, Q, R and S are real quantities. Collating equations (18) and (25), we see that

$$\mu/\pi = \frac{QR-PS}{PR+QS} \quad \text{and} \quad (c_y/c)^2 = \frac{PR+QS}{R^2+S^2}. \qquad \ldots\ldots(26)$$

These four sets of equations (23) to (26) provide the necessary link between the parameters μ and c, which are derived from ultrasonic measurements, and the parameters ΔH_m, ΔE_m, C_{1m} and τ_m, which describe a molecular process. The absorption μ must be sufficiently small to permit the use of standard thermodynamic formulae and to satisfy the approximation made between equations (3) and (4), namely

$$(\alpha c/\omega)^2 \ll 1. \qquad \ldots\ldots(27)$$

2.6. *An Isolated Relaxation Mechanism*

Increasing the frequency of a sound wave propagated in a liquid leads to the desistance of successive internal degrees of freedom, and a consequent decrease in the effective specific heat; at sufficiently high frequencies only the residual specific heat $C_{\mathrm{p}y}$, $C_{\mathrm{v}y}$ remains.

The static specific heat $C_{\mathrm{p}0}$, $C_{\mathrm{v}0}$ is a sum of components due to vibration, rotation and other internal causes, both thermal and structural, plus the residual, or infinite frequency, specific heat. In general, therefore, there will be several terms in the summations in equation (23). However, if measurements of the ultrasonic absorption are made over a small frequency range, it is reasonable to suppose that only one of the relaxing components of the specific heat, $C_{\mathrm{p}m} = C_{\mathrm{p}a}$, has a relaxation time τ_a such that $\omega\tau_a \simeq 1$. The components of specific heat now fall into three categories:

$$\left.\begin{array}{lll} C_{\mathrm{p}m} = C_{\mathrm{p}m}' & \text{for which} & \omega\tau_m = \omega\tau_m' \ll 1 \\ C_{\mathrm{p}m} = C_{\mathrm{p}a} & \text{for which} & \omega\tau_m = \omega\tau_a \simeq 1 \\ \text{and the remaining } C_{\mathrm{p}m} & \text{for which} & \omega\tau_m \gg 1 \end{array}\right\} \ldots\ldots(28)$$

The use of a prime in the following denotes that quantities are related to $C_{\mathrm{p}m}'$ and τ_m' ($\ll \tau_a$).

By defining the parameters

$$K = 1 + K_a + 2\sum_m K_{ma}'C_{1m}'$$

$$\left.\begin{array}{ll} C_{\mathrm{p}1} = C_{\mathrm{p}y} + \sum\limits_m C_{1m}'; & C_{\mathrm{p}1}{}^0 = C_{\mathrm{p}1} + C_{1a} \\ C_{\mathrm{p}k} = C_{\mathrm{p}y} + \sum\limits_m (1 + K_m' + \sum\limits_n K_{mn}'C_{1n}')\, C_{1m}'; & C_{\mathrm{p}k}{}^0 = C_{\mathrm{p}k} + KC_{1a} \\ J = \sum\limits_m \tau_m' C_{1m}'; & J^0 = J + \tau_a C_{1a} \\ J_k = \sum\limits_m (1 + K_m' + 2\sum\limits_n K_{mn}'C_{1n}')\, \tau_m' C_{1m}'; & J_k{}^0 = J_k + K\tau_a C_{1a} \end{array}\right\} \ldots\ldots(29)$$

and introducing the conditions (28) into equations (23) to (26), it may be shown that

$$\mu/f = \alpha c/f^2 = \frac{A}{1+(f/f_a)^2} + B \qquad \ldots\ldots(30)$$

and

$$(c/c_0)^2 = 1 + \frac{(KC_{p1}{}^0 - C_{pk}{}^0)/C_{p1}{}^0 C_{pk}}{1+(f_a/f)^2} C_{ia}. \qquad \ldots\ldots(31)$$

Here c_0 is the limiting velocity as ω tends to zero (lower frequency relaxing processes being temporarily removed), and

$$f_a = \left[\frac{C_{p1}{}^0 C_{pk}{}^0}{C_{p1} C_{pk}}\right]^{1/2} \Big/ 2\pi\tau_a; \quad B = 2\pi^2\left(\frac{J_k}{C_{pk}} - \frac{J}{C_{p1}}\right); \quad A+B = 2\pi^2\left(\frac{J_k{}^0}{C_{pk}{}^0} - \frac{J^0}{C_{p1}{}^0}\right). \qquad \ldots\ldots(32)$$

Equation (30) may be written

$$\mu = \alpha c/f = \frac{2\mu_a(f/f_a)}{1+(f/f_a)^2} + Bf, \qquad \ldots\ldots(33)$$

where $\mu_a = \frac{1}{2}Af_a$ is the maximum value of the first (relaxing) term on the right-hand side of (33). A useful expression for μ_a is

$$\frac{2\mu_a}{\pi} = \frac{Af_a}{\pi} = \frac{KC_{p1}{}^0 - C_{pk}{}^0}{(C_{p1}{}^0 C_{pk}{}^0 C_{p1} C_{pk})^{1/2}} C_{ia} - 2\pi f_a \left[\frac{KJ_k}{C_{pk}{}^0 C_{pk}} - \frac{J}{C_{p1}{}^0 C_{p1}}\right] C_{ia}. \qquad \ldots\ldots(34)$$

In the case of a gas $\Delta E_m = \Delta H_m$ for all values of m, and then

$$\frac{2\mu_a}{\pi} = \frac{(C_{p1}{}^0 - C_{v1}{}^0)}{(C_{p1}{}^0 C_{v1}{}^0 C_{p1} C_{v1})^{1/2}} C_{1a} - \frac{Bf_a}{\pi} \frac{(C_{p1} + C_{v1}{}^0)}{C_{p1}{}^0 C_{v1}{}^0} C_{1a}. \qquad \ldots\ldots(35)$$

The first terms on the right-hand sides of equations (34) and (35) become identical when C_{1a} is the specific heat of a thermal process. This is proved by introducing the conditions (28) into equations (30) and (32). The second terms on the right-hand sides of equations (34) and (35) are correction terms arising from the effect of higher frequency dispersions.

Equation (35) may be deduced from equation (4) by replacing the adiabatic compressibility β_ω by $\beta_{T\omega}/\gamma_\omega$. The isothermal compressibility $\beta_{T\omega}$ is then assumed to be frequency-independent, and the ratio of the specific heats is written

$$\gamma_\omega = C_{p\omega}/C_{v\omega} = \frac{C_{py} + \sum_m C_{im}/(1+j\omega\tau_m)}{C_{vy} + \sum_m C_{im}/(1+j\omega\tau_m)}. \qquad \ldots\ldots(36)$$

Separating real and imaginary parts of the resulting equation, and introducing the conditions (28), we find a relation identical to (35). This procedure is similar to the classical treatments of Einstein (1920), Kneser (1931), Herzfeld (1941) and others. The restriction on the isothermal compressibility eliminates all structural processes. In the present analysis this restriction has been removed by considering the adiabatic compressibility as a whole, and general formulae have been deduced which include both thermal and structural processes.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Willis Jackson for his interest in this work, and they are grateful to Dr. R. O. Davies for helpful suggestions.

APPENDIX

I. *The Derivation of Equations* (16) *and* (17)

The derivation of equation (16) starts from the formula

$$\theta_\infty = V^{-1}(\partial V/\partial T)_{P,\infty} = -V^{-1}(\partial S/\partial P)_{T,\infty} \qquad \ldots\ldots(\mathrm{A}\,1)$$

which provides alternative expressions for the thermal expansion coefficient θ_∞. Expand each of the derivatives $(\partial V/\partial T)_{P,\infty}$ and $(\partial S/\partial P)_{T,\infty}$ according to formula (9), and equate corresponding terms of the two forms of the expansion

$$\theta_\infty = \theta_y + \sum_m \theta_m\,;$$

it is found that

$$(\partial y_m/\partial P)_{T,\infty} = -\frac{T\Delta V_m}{(\Delta H_m)^2}\,C_{\mathrm{p}m}\,. \qquad \ldots\ldots(\mathrm{A}\,2)$$

The derivative $T(\partial P/\partial T)_{S,\infty}$ and $P_\infty{}^*$ may be written in terms of θ by means of

$$T(\partial P/\partial T)_{S,\infty} = \gamma_\infty P_\infty{}^*/(\gamma_\infty - 1) = \theta_\infty T/\beta_\infty(\gamma_\infty - 1) = C_{\mathrm{p}\infty}/\theta_\infty V. \qquad \ldots\ldots(\mathrm{A}\,3)$$

The two derivatives in the expression (11) for β_m may now be deduced:

$$-V^{-1}(\partial V/\partial y_m)_{S,P,m} = -V^{-1}(\partial V/\partial y_m)_{P,T,m} + V^{-1}(\partial V/\partial S)_{P,y}(\partial S/\partial y_m)_{P,T,m}$$

$$= \frac{\theta_y \Delta H_m}{C_{\mathrm{p}y}}\left(1 - \frac{\Delta V_m}{\Delta H_m}\frac{C_{\mathrm{p}y}}{V\theta_y}\right), \qquad \ldots\ldots(\mathrm{A}\,4)$$

$$(\partial y_m/\partial P)_{S,\infty} = (\partial y_m/\partial P)_{T,\infty} + (\partial y_m/\partial T)_{P,\infty}(\partial T/\partial P)_{S,\infty}$$

$$= \frac{\theta_y VT}{\Delta H_m}\frac{C_{\mathrm{p}m}}{C_{\mathrm{p}\infty}}\left[1 - \frac{C_{\mathrm{p}y}}{V\theta_y}\frac{\Delta V_m}{\Delta H_m} + \sum_n \frac{C_{\mathrm{p}n}}{V\theta_y}\left(\frac{\Delta V_n}{\Delta H_n} - \frac{\Delta V_m}{\Delta H_m}\right)\right]. \qquad \ldots\ldots(\mathrm{A}\,5)$$

The product of (A 4) and (A 5) leads to

$$\beta_m/\beta_y = \left(\gamma_y \frac{\Delta E_m}{\Delta H_m} - 1\right)\frac{C_{\mathrm{p}m}}{C_{\mathrm{p}\infty}}\left[\frac{\gamma_y \Delta E_m/\Delta H_m - 1}{\gamma_y - 1} + \sum_n \frac{C_{\mathrm{p}n}}{\theta_y V}\left(\frac{\Delta V_n}{\Delta H_n} - \frac{\Delta V_m}{\Delta H_m}\right)\right]. \qquad \ldots\ldots(\mathrm{A}\,6)$$

Eliminating θ_y by means of (A 3) and carrying out the summation

$$\beta_\infty/\beta_y = 1 + \sum_m \beta_m/\beta_y,$$

we obtain equation (16).

The derivation of equation (17) follows two paths. Starting from an expansion of the isothermal compressibility, $\beta_{T\infty} = -V^{-1}(\partial V/\partial P)_{T,\infty}$, and recognizing the relation $\beta_{T\infty} = \gamma_\infty \beta_\infty$, it may be shown, by comparison with equation (16), that

$$\sum_m\left(C_{\mathrm{v}m} - \frac{\Delta E_m}{\Delta H_m}C_{\mathrm{p}m}\right)\sum_n\left(C_{\mathrm{v}n} - \frac{\Delta E_n}{\Delta H_n}C_{\mathrm{p}n}\right)\Big| = (C_{\mathrm{p}\infty} - C_{\mathrm{v}\infty})\sum_m\left(C_{\mathrm{p}m}\left(\frac{\Delta E_m}{\Delta H_m}\right)^2 - C_{\mathrm{v}m}\right) \qquad \ldots\ldots(\mathrm{A}\,7)$$

On the other hand, $C_{\mathrm{p}\omega} - C_{\mathrm{v}\omega}$ may be expanded with θ_ω and P_ω^* in the formula $C_{\mathrm{p}\omega} - C_{\mathrm{v}\omega} = \theta_\omega V P_\omega^*$, the result being

$$\sum_m \frac{\Delta H_m}{\Delta E_m}\left(C_{\mathrm{v}m} - \frac{\Delta E_m}{\Delta H_m} C_{\mathrm{p}m}\right)\sum_n \left(C_{\mathrm{v}n} - \frac{\Delta E_n}{\Delta H_n} C_{\mathrm{p}n}\right)$$
$$= (C_{\mathrm{p}\omega} - C_{\mathrm{v}\omega}) \sum_m \frac{\Delta H_m}{\Delta E_m}\left(C_{\mathrm{p}m}\left(\frac{\Delta E_m}{\Delta H_m}\right)^2 - C'_{\mathrm{v}m}\right). \qquad \text{......(A 8)}$$

Since both of the equations (A 7) and (A 8) must be true for all frequencies $\omega/2\pi$ equation (17) would appear to be the necessary condition of compatibility.

II. *Equation* (16) *with One Relaxing Process Only*

When only one process needs to be considered, equation (16) reduces to

$$\frac{\beta_\omega}{\beta_y} = 1 + \frac{C_{\mathrm{p}m}}{(\gamma_y - 1)C_{\mathrm{p}\omega}}\left(\gamma_y \frac{\Delta E_m}{\Delta H_m} - 1\right)^2. \qquad \text{......(A 9)}$$

Referring to zero frequency ($\omega = 0$), and denoting the total relaxing components of the various parameters by

$$\Delta\beta = \beta_0 - \beta_y, \qquad \Delta C_{\mathrm{p}} = C_{\mathrm{p}0} - C_{\mathrm{p}y},$$
$$\Delta\theta = \theta_0 - \theta_y, \qquad \Delta C_{\mathrm{v}} = C_{\mathrm{v}0} - C_{\mathrm{v}y},$$

we can rewrite (A 9)

$$\frac{\Delta\beta}{\beta_y} = \frac{\Delta C_{\mathrm{p}}}{(\gamma_y - 1)C_{\mathrm{p}0}}\left(\gamma_y \frac{\Delta E_m}{\Delta H_m} - 1\right)^2. \qquad \text{......(A 10)}$$

It may be shown that this is identical to

$$\frac{\Delta\beta}{\beta_0} = \frac{C_{\mathrm{p}0} - C_{\mathrm{v}0}}{C_{\mathrm{v}0}}\left(\frac{\Delta\theta}{\theta_0} - \frac{\Delta C_{\mathrm{p}}}{C_{\mathrm{p}0}}\right)^2 \Big/ \left\{\frac{\Delta C_{\mathrm{p}}}{C_{\mathrm{p}0}}\left(1 - \frac{\Delta C_{\mathrm{p}}}{C_{\mathrm{p}0}}\right)\right\}, \qquad \text{......(A 11)}$$

a relation which has been derived by Davies (1954) in a different manner.

REFERENCES

DAVIES, R. O., 1954, *Proc. Roy. Soc.* A, **226**, 24.
DAVIES, R. O., and LAMB, J., 1956, *Proc. Phys., Soc.* B, **69**, 293.
EINSTEIN, A., 1920, *S. B. Preuss. Akad. Wiss.*, p. 380.
HALL, L., 1948, *Phys. Rev.*, **73**, 775.
HERZFELD, K. F., 1941, *J. Acoust. Soc. Amer.*, **13**, 33.
KNESER, H. O., 1931, *Ann. Phys., Lpz.*, **11**, 761.
MEIXNER, J., 1952, *Acustica*, **2**, 101.

Phase Distribution near Focus in an Aberration-free Diffraction Image

By E. H. LINFOOT[†] and E. WOLF[‡]
[†]The Observatories, University of Cambridge
[‡]The Physical Laboratories University of Manchester

MS. received 12*th October* 1955, *and in revised form* 24*th February* 1956

Abstract. A satisfactory picture of the three-dimensional intensity distribution near focus in the aberration-free diffraction image of a monochromatic point object was first given by Zernike and Nijboer in 1949. No corresponding general picture of the phase relations between different parts of the image seems to have been worked out, although Gouy's discovery in 1890 of the so-called phase anomaly at focus aroused an interest in this topic which, as a succession of publications shows, has not decreased in the intervening 66 years.

In the present paper a sharpened version of Lommel's classical analysis is applied to obtain a general picture of the phase distribution near focus and to examine in detail, in the special case of an F/3·5 pencil, its peculiarities near the geometrical focal point, the Airy dark rings and the axial nodes (points of zero intensity) of the diffraction image. The 'phase anomaly' near focus and the singular behaviour of the phase along the axis of the pencil become more readily intelligible when they are considered against the background of this general picture.

§ 1. Introduction

A central problem in the theory of image formation in optical instruments is the determination of the three-dimensional light distribution which represents the aberration-free image of a point source by an axially symmetric optical system. Analytical formulae for the intensity distribution in such an image were first given by von Lommel (1885) and almost simultaneously by Struve (1886). The first satisfactory diagram of the intensity distribution was published by Zernike and Nijboer (1949) more than sixty years later,[†] and similar diagrams for images affected by selected amounts of spherical aberration have been worked out by them (1949) and by A. Maréchal and his co-workers (1948).

Less progress seems to have been made in the complementary part of the problem, namely the study of the phase distribution in the three-dimensional diffraction image, even though modern developments in microwave optics have greatly enhanced its practical interest. Theoretical interest was never lacking; as early as 1890 Gouy discovered the so-called phase anomaly near focus, and the subject has since been treated by many authors[‡], notably by Joubin (1892), Fabry (1893), Julius (1895), Zeeman (1897, 1900, 1901), Sagnac (1903, 1904 a, b), Debye (1909), Reiche (1909 a, b), Ignatowsky (1919), Fokker (1923, 1924), Picht (1930), Rubinowicz (1938), Breuninger (1938, 1939), Bouwkamp (1940) and Toraldo di Francia (1942).

[†] An inaccurate diagram, widely reproduced in optical textbooks, was given by Berek (1926). The isophotes in figure 2 were constructed from tables given by von Lommel (1885).

[‡] For a review of the early papers, see Reiche (1909 a).

All these investigations were concerned with the aberration-free image and their joint outcome was, broadly speaking, to confirm Gouy's results, to determine the detailed behaviour of the phase distribution along particular rays through the geometrical focus, and to establish that the properties of the phase distribution along the axial ray are qualitatively different from those elsewhere. The last result seems to have been first established by Picht (1930) However, a satisfactory overall picture of the three-dimensional phase distribution, as distinct from the phase distribution along isolated rays, did not emerge.

Such a picture seems nowadays hardly less essential to a proper understanding of the diffraction image than that obtained by Zernike and Nijboer for the three-dimensional intensity distribution. In the present paper we go some way towards obtaining one by an application of Lommel's classical formulae which, for pencils of not too large numerical aperture, allow the calculation of the phases at points not too far from the geometrical focus. From the computed values of the phases at a sufficiently dense set of sample points, the phase distribution near the geometrical focus has been obtained by graphical methods.

It turns out that in the geometrical pencil of rays near focus the co-phasal surfaces are very nearly plane; the corresponding phenomenon for sound waves can be seen in schlieren photographs given by Pohl (1948). The singular behaviour of the phase along the axis becomes readily intelligible when its connection with the existence of points of zero intensity is made visible to the eye (see figures 5 and 6). The singular behaviour of the phase at those points of the geometrical focal plane which corresponds to the Airy dark rings becomes intelligible in the same way (figures 3 and 4).

§ 2. PHASE NEAR FOCUS

The notation and approximations are similar to those of our paper (1952) on telescopic star images. In figure 1, ABA′B′ represents a circular aperture through which issues a train of converging spherical waves of wavelength λ. The diameter AA′ of this aperture is of length $2a$; C is the pole of the wave surface S which momentarily fills it. O is the centre of curvature of S, $CO=f$ its radius of curvature. It is assumed throughout that $a^2 \ll f^2$; for example, $a^2/f^2=1/49$ in an F/3·5 pencil. P (x, y, z) is an arbitrary point in the space near O.

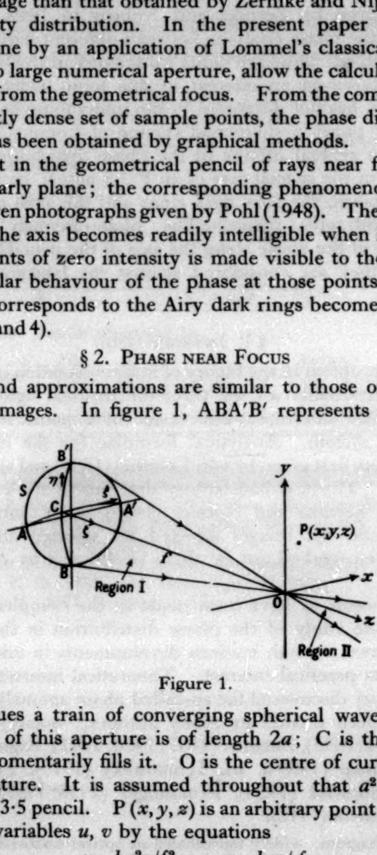


Figure 1.

We define the variables u, v by the equations

$$u=ka^2z/f^2, \qquad v=kar/f, \qquad \ldots\ldots(2.1)$$

where $k=2\pi/\lambda$ and r stands for $\sqrt{(x^2+y^2)}$. In physical terms, $u/4\pi$ is the number of fringes of defocusing and v/π the number of fringes of lateral displacement of P relative to O. We note that $|v/u|\lessgtr 1$ according as P lies in the geometrical cone of rays or in the geometrical shadow.

The complex displacement† at $P(x, y, z)$ in the space near O which results from waves of unit amplitude in S is given by the approximate formula

$$D_\lambda(\mathrm{P})=\frac{2\pi i}{\lambda}a^2\exp[ik(f-R')]\int_0^1\exp(\tfrac{1}{2}iu\rho^2)\mathrm{J}_0(v\rho)\rho\,d\rho, \qquad \ldots\ldots(2.2)$$

where R' denotes the distance CP and J_0 is the Bessel function of order zero.

The equation (2.2) is valid, broadly speaking, at points P for which the number of fringes of defocusing and of lateral displacement are both O(1), that is to say do not exceed about 5 or 10. To define its domain of validity more precisely, we should need to specify the amount of inaccuracy which is regarded as tolerable and to examine the approximation errors in the same way as was done, for an F/15 pencil, in our paper on the intensity distribution already referred to.

The integral on the right of (2.2) can be evaluated in terms of the functions

$$U_n(u,v)=\sum_{m=0}^{\infty}(-1)^m\left(\frac{u}{v}\right)^{n+2m}\mathrm{J}_{n+2m}(v) \qquad \ldots\ldots(2.3)$$

introduced by von Lommel for the purpose; in fact‡

$$\int_0^1\mathrm{J}_0(v\rho)\exp(\tfrac{1}{2}iu\rho^2)\rho\,d\rho=-iu^{-1}\exp(\tfrac{1}{2}iu)(U_2+iU_1) \qquad \ldots\ldots(2.4)$$

and (2.2) therefore gives

$$D_\lambda(\mathrm{P})=\frac{2\pi a^2}{\lambda f u}\exp\{i[k(f-R')+\omega+\tfrac{1}{2}u]\}\surd(U_1^2+U_2^2), \qquad \ldots\ldots(2.5)$$

where

$$\cos\omega=\frac{U_2}{\surd(U_1^2+U_2^2)},\quad \sin\omega=\frac{U_1}{\surd(U_1^2+U_2^2)} \qquad \ldots\ldots(2.6)$$

and $\surd$ denotes the positive square root here and elsewhere.

In equation (2.5)

$$\begin{aligned}R'=\mathrm{CP}&=\surd[x^2+y^2+(z+f)^2]\\&=(z+f)\left[1+\frac{r^2}{(z+f)^2}\right]^{1/2}\\&=(z+f)+\frac{1}{2}\frac{r^2}{z+f}-\frac{1}{8}\frac{r^4}{(z+f)^3}-\ldots \qquad \ldots\ldots(2.7)\end{aligned}$$

Hence, by (2.1),

$$\begin{aligned}k(f-R')&=-\left(\frac{f}{a}\right)^2u-\mathrm{O}\left(\frac{kr^2}{2f}\right)\\&=-\left(\frac{f}{a}\right)^2u-\frac{\lambda}{4\pi f}\left(\frac{f}{a}\right)^2\mathrm{O}(v^2). \qquad \ldots\ldots(2.8)\end{aligned}$$

In the region of space where

$$u=\mathrm{O}(1),\quad v=\mathrm{O}(1), \qquad \ldots\ldots(2.9)$$

the error term on the right of (2.8) is negligible in comparison with unity in a practical case; for example, if $\lambda=2\times10^{-5}$ inch the value of $\lambda f/4\pi a^2$ is approximately $F^2/157\,000$, where F is the focal ratio and f is the distance CO measured in inches. Thus we may safely use the approximation

---

† Following Rayleigh, we give this name to the time-independent part $D_\lambda(\mathrm{P})$ of the wave displacement D (P) exp ($ikct$).

‡ See Watson 1922, p. 541.

$$k(f-R') = -\left(\frac{f}{a}\right)^2 u. \qquad \ldots\ldots(2.10)$$

It follows from (2.5) that, in the region where (2.9) holds, the amplitude $M(u, v)$ and the phase $\phi(u, v)$ of the complex displacement $D_\lambda(\mathrm{P})$ are given by the equations

$$M(u,v) = \frac{2\pi a^2}{\lambda f|u|}\sqrt{(U_1{}^2 + U_2{}^2)}, \qquad \ldots\ldots(2.11)$$

$$\phi(u,v) = \omega + \tfrac{1}{2}u - (f/a)^2 u \,(\mathrm{mod}\, 2\pi) \qquad \ldots\ldots(2.12)$$

respectively.†

From (2.12) and (2.6) it appears that $\phi(u, v)$, unlike $M(u, v)$, cannot be expressed in terms of the parameters u, v alone but has a structure which changes as the focal ratio varies. The multivalued function $\phi(u, v)$ has a branch point at each zero of the intensity $M^2(u, v)$ and is continuous elsewhere. At the focal point $u = v = 0$ (which is not a branch point) one of its values is $\pi/2$.

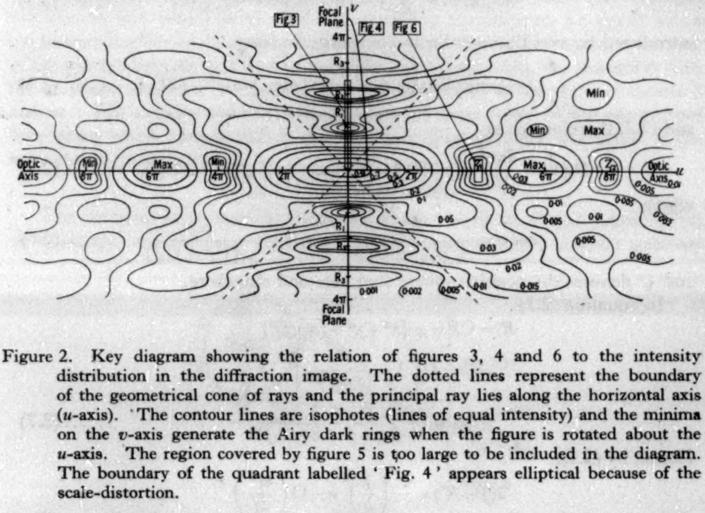


Figure 2. Key diagram showing the relation of figures 3, 4 and 6 to the intensity distribution in the diffraction image. The dotted lines represent the boundary of the geometrical cone of rays and the principal ray lies along the horizontal axis (u-axis). The contour lines are isophotes (lines of equal intensity) and the minima on the v-axis generate the Airy dark rings when the figure is rotated about the u-axis. The region covered by figure 5 is too large to be included in the diagram. The boundary of the quadrant labelled 'Fig. 4' appears elliptical because of the scale-distortion.

The co-phasal surfaces $\phi(u, v) = $ constant are surfaces of revolution about the axis $v = 0$. By (2.12), (2.6), (2.3) they possess the further symmetry expressed by the equation

$$\phi(-u, v) + \phi(u, v) = \pi \;(\mathrm{mod}\, 2\pi). \qquad \ldots\ldots(2.13)$$

The form of the equiphase surfaces can be computed from (2.12) and (2.6). We first consider the region, very close to the geometrical focal plane, represented in figure 2 by the rectangle labelled 'Fig. 3'. This region is a thin, disc-shaped volume extending a little beyond the second Airy dark ring R_2. Close to the

---

† The notation (mod 2π) means that the two sides of the equation differ either by zero or by a multiple of 2π. ϕ itself is indeterminate to the extent of an additive multiple of 2π.

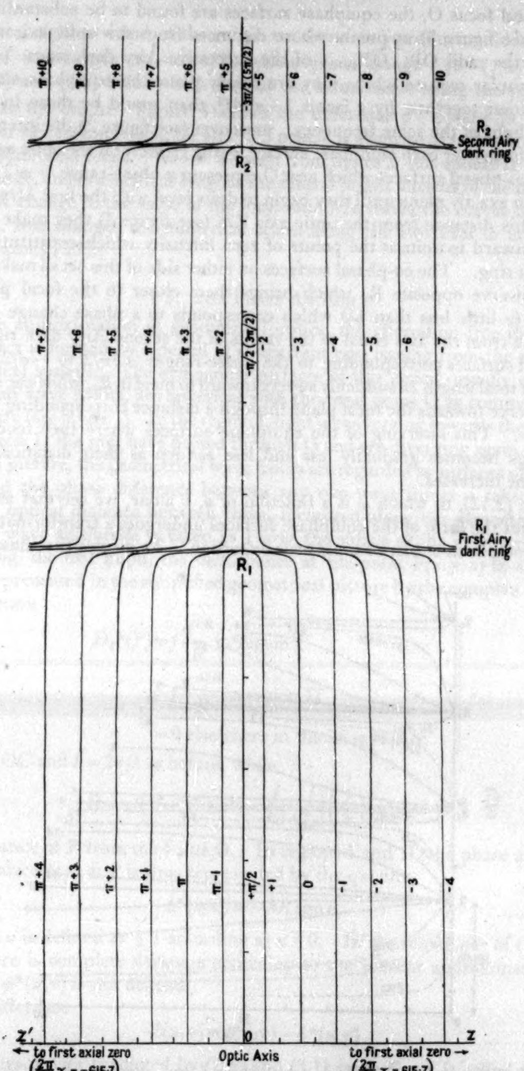

π+10
π+9
π+8
π+7
π+6
π+5
-3π/2 (5π/2)
-5
-6
-7
-8
-9
-10
R₂
Second Airy dark ring
R₂
π+7
π+6
π+5
π+4
π+3
π+2
-π/2 (3π/2)
-2
-3
-4
-5
-6
-7
R₁
First Airy dark ring
R₁
π+4
π+3
π+2
π+1
π
π-1
+π/2
+1
0
-1
-2
-3
-4
Z'
0
Z
← to first axial zero
(2π/λ z = -615·7)
Optic Axis
to first axial zero →
(2π/λ z = 615·7)

Figure 3. Co-phasal surfaces in the neighbourhood of the geometrical focal planeofan F/3·5 pencil. O is the geometrical focus, ZZ′ the axis of the pencil, OR_1 and OR_2 the radii of the first and second Airy dark rings.

geometrical focus O, the equiphase surfaces are found to be substantially plane except (see figure 3) at points whose distances from the optic axis are nearly equal to the radii OR_1, $OR_2, \ldots$ of the successive Airy dark rings. However, in the annular regions where they are nearly plane the equiphase surfaces are spaced closer together, by a factor $1-a^2/4f^2$, than would be those in a parallel beam of light of the same frequency; moreover (see figure 2) the intensity is far from uniform over each equiphase surface, being greatest on the optic axis.

The co-phasal surfaces which near O represent a phase-range $-\pi/2$ to $+3\pi/2$ are almost exactly plane until they begin to draw level with the first Airy dark ring R_1; at this distance from the optic axis ZZ′ (see figure 3) they make a sudden swerve inward to unite at the points of zero intensity which constitute the first Airy dark ring. The co-phasal surfaces on either side of this set all make a similar inward swerve opposite R_1 which brings them closer to the focal plane by a distance (a little less than $\frac{1}{2}\lambda$) which corresponds to a phase change of π. At a distance from the axis equal to the radius of the second Airy dark ring R_2, the co-phasal surfaces corresponding to the phase-ranges $-3\pi/2$ to $-\pi/2$ and $3\pi/2$ to $5\pi/2$ of total length 2π suddenly swerve inward to meet in R_2, while the remainder again swerve towards the focal plane through a distance corresponding to a phase change π. This swerving of the equiphase surfaces where they face the Airy dark rings becomes gradually less and less sudden as their distance from the focal plane increases.

From (2.12), in which ω is a function of u, v alone, we see that as the focal ratio varies the form of the equiphase surfaces undergoes a transformation which is more complicated than a simple scale distortion. To display the phase distribu-

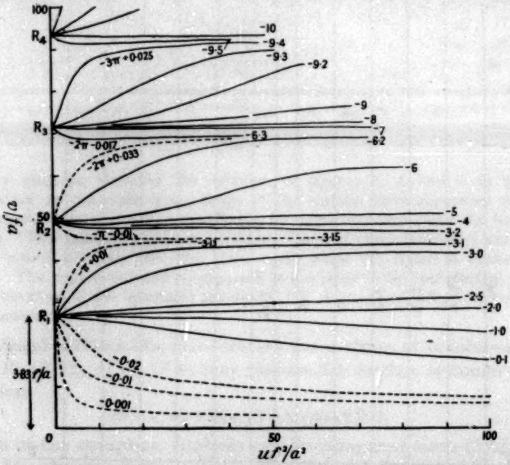


Figure 4. Difference $\phi-\tilde{\phi}$ between actual phase ϕ and phase $\phi=\frac{1}{4}u-(f^2/a^2)u$ of the fiducial plane wave. The numerical scales refer to an F/3·5 pencil; in this case, unit distance on the graph represents an actual distance $\lambda/2\pi$, the radius OR_1 of the first Airy dark ring is $3{\cdot}83f/a$ units $=0{\cdot}61\lambda f/a$, and the distance OZ_1 to the first axial node Z_1 (not shown) is $4\pi f^2/a^2=606$ units, nearly.

tion in the neighbourhood of the origin we therefore introduce a fiducial plane wave,† of wavelength $\lambda(1-a^2/4f^2)$, whose phase $\tilde{\phi}=\frac{1}{2}\pi+\frac{1}{4}u-(f/a)^2u$ agrees (see (2.12) and (3.8) below) with that of the actual wave along the part of the u-axis near the origin. Figure 4 shows the difference $\phi-\tilde{\phi}=\omega+\frac{1}{4}u-\frac{1}{2}\pi$ over the interior of a region which, for an F/3·5 pencil, is a sphere of radius 16λ centred on the geometrical focus. As will be seen from figure 2, this sphere is about 3·5 times as wide, and one-sixth as long, as the central bright nucleus of the diffraction image. A change in the value of the focal ratio $f/2a$ leaves the curves in figure 4 unaltered, but changes the numerical scales along the two axes in accordance with equations (2.1).

§ 3. THE PHASE ANOMALY

In the approximation of geometrical optics, the aberration-free image of a point object is formed by a pencil of concurrent rays issuing from the exit pupil of an optical system. The orthogonal surfaces of this pencil, usually called the geometrical wave fronts, are spherical with the focal point O as common centre of curvature (see figure 1). In the region I, light is converging towards the focus O; in the region II the rays have passed through O and are diverging again. In this simplified picture, the geometrical wave fronts are regarded as surfaces of constant phase, and the phase difference between any two wave fronts is taken as $2\pi/\lambda$ times the optical distance between them, measured along the geometrical rays.

If the wave amplitude is taken as 1 over the suface of the geometrical wave front filling the exit pupil, the disturbance at the point $P(x,y,z)$ in the image space is represented in the simplified geometrical picture by the complex displacement function

$$\left.\begin{aligned} D_\lambda{}^*(\mathrm{P}) &= f\frac{\mathrm{e}^{ikR}}{R} \text{ in region I} \\ &= f\frac{\mathrm{e}^{-ikR}}{R} \text{ in region II} \\ &= 0 \text{ elsewhere in the image space} \end{aligned}\right\}, \qquad \ldots\ldots(3.1)$$

where $f=\mathrm{OC}$ and $k=2\pi/\lambda$ as before, while

$$R=\sqrt{(x^2+y^2+z^2)}=\frac{f}{ka}\sqrt{\left\{\left(\frac{f}{a}\right)^2u^2+v^2\right\}}$$

is the distance of P from the focus O. In regions I and II, the phase attributed to the displacement at P is thus represented by the quantity

$$\phi^*(u,v)=-kR\,\mathrm{sgn}\,u, \qquad \ldots\ldots(3.2)$$

where $\mathrm{sgn}\,u$ is defined as ± 1 according as $u \gtrless 0$. In the remainder of the image space, there is complete darkness according to the present approximation, and the phase $\phi^*(u,v)$ is not defined.

The difference

$$\delta(u,v)=\phi(u,v)-\phi^*(u,v) \qquad \ldots\ldots(3.3)$$

between the phases predicted by (2.5) and (3.1) respectively is called the *phase anomaly*. It is defined only in the regions I and II, where by (3.2)

$$\delta(u,v)=\phi(u,v)+kR\,\mathrm{sgn}\,u \qquad \ldots\ldots(3.4)$$

---

† This expedient was suggested by a referee.

and by (2.13) and (3.4)

$$\delta(-u, v) + \delta(u, v) = \pi \pmod{2\pi}. \qquad \ldots\ldots(3.5)$$

When P tends to O, the geometrical phase $\phi^*(0, 0)$ tends to zero, while the 'principal value' of $\phi(0, 0)$ is $\frac{1}{2}\pi$ by (2.2); (3.4) then gives the well-known result

$$\delta(0, 0) = \tfrac{1}{2}\pi. \qquad \ldots\ldots(3.6)$$

In the parts of the regions I and II which lie within 6λ of the geometrical focus, each surface of phase ϕ is substantially coincident with the 'best' plane approximation to the corresponding geometrical wave cap of phase $\phi - \frac{1}{2}\pi$, namely the plane which deviates from the geometrical wave cap by equal and opposite amounts at centre and edge. This has the consequence, already noted above, that consecutive surfaces of the same phase (mod 2π) in this region are spaced a distance $(1 - a^2/4f^2)\lambda$ apart, in agreement with (2.12) and (3.9), instead of a distance λ apart.

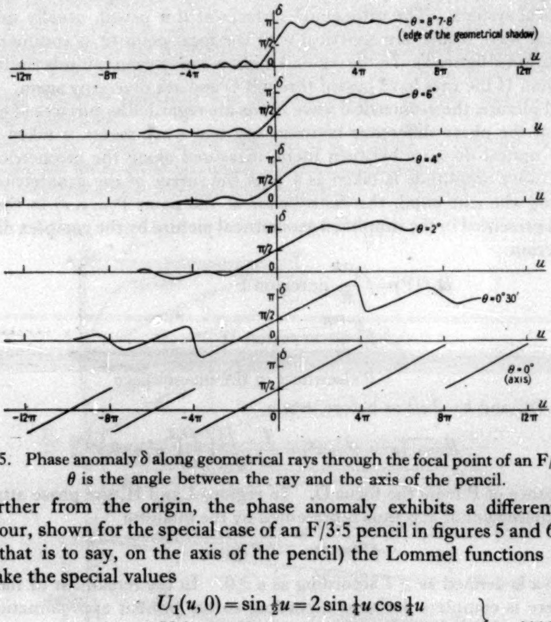


Figure 5. Phase anomaly δ along geometrical rays through the focal point of an F/3·5 pencil. θ is the angle between the ray and the axis of the pencil.

Further from the origin, the phase anomaly exhibits a different type of behaviour, shown for the special case of an F/3·5 pencil in figures 5 and 6. When $v = 0$ (that is to say, on the axis of the pencil) the Lommel functions U_1, U_2 in (2.6) take the special values

$$\left.\begin{aligned} U_1(u, 0) &= \sin \tfrac{1}{2}u = 2 \sin \tfrac{1}{4}u \cos \tfrac{1}{4}u \\ U_2(u, 0) &= 1 - \cos \tfrac{1}{2}u = 2 \sin^2 \tfrac{1}{4}u \end{aligned}\right\} \qquad \ldots\ldots(3.7)$$

and we obtain from (2.6)

$$\left.\begin{aligned} \omega &= \tfrac{1}{2}\pi - \tfrac{1}{4}u \pmod{2\pi} \quad \text{if} \quad 2s\pi < \tfrac{1}{4}u < (2s+1)\pi \\ &= \tfrac{3}{2}\pi - \tfrac{1}{4}u \pmod{2\pi} \quad \text{if} \quad (2s+1)\pi < \tfrac{1}{4}u < 2s\pi \end{aligned}\right\} \qquad \ldots\ldots(3.8)$$

where $s = 0, \pm 1, \pm 2, \ldots$. Hence by (2.12), (3.1), (3.3)

$$\begin{aligned} \delta(u, 0) &= \tfrac{1}{2}\pi + \tfrac{1}{4}u \pmod{2\pi} \quad \text{if} \quad 8s\pi \leqslant |u| < 8(s+\tfrac{1}{2})\pi \\ &= -\tfrac{1}{2}\pi + \tfrac{1}{4}u \pmod{2\pi} \quad \text{if} \quad 8(s+\tfrac{1}{2})\pi \leqslant |u| < 8(s+1)\pi. \end{aligned} \qquad \ldots\ldots(3.9)$$

There is a phase jump of π at each of the axial 'nodes' (points of zero intensity) $u=4s\pi$; $s=\pm 1, \pm 2, \ldots$. Between two consecutive nodes the phase anomaly $\delta(u, 0)$ varies linearly, as shown in the bottom curve of figure 5, until $|u|$ becomes so large that the approximation (2.2) begins to break down.

The computation of the phase distribution on the edge $v=|u|$ of the geometrical shadow is simplified by observing that the Lommel functions U_1, U_2 take for $v=u$ the simple forms†

$$U_1(u, u)=\tfrac{1}{2}\sin u$$
$$U_2(u, u)=\tfrac{1}{2}[\mathrm{J}_0(u)-\cos u]. \qquad \ldots\ldots(3.10)$$

The resulting phase anomaly is shown, for the case of an F/3·5 pencil, in the top curve of figure 5.

Figure 5 illustrates the difference between the behaviour of the phase anomaly along the axial ray and along the non-axial rays of the pencil. Along each non-axial ray, the phase anomaly approaches π as the distance from the focus increases (Debye 1909). The different behaviour, described by (3.9), of the phase anomaly along the axial ray is by no means surprising, since this ray passes through the axial nodes. In figure 6 is shown the behaviour of the co-phasal surfaces near the first axial node.

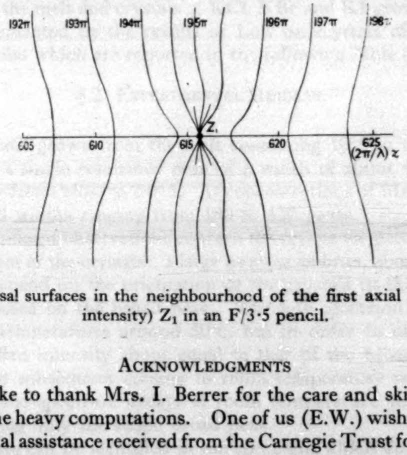


Figure 6. Co-phasal surfaces in the neighbourhood of the first axial node (point of zero intensity) Z_1 in an F/3·5 pencil.

Acknowledgments

We should like to thank Mrs. I. Berrer for the care and skill with which she has carried out the heavy computations. One of us (E.W.) wishes to acknowledge gratefully financial assistance received from the Carnegie Trust for the Universities of Scotland towards computing expenses. He is also indebted to the Universities of Edinburgh and Manchester for awards of I.C.I. Research Fellowships.

References

Berek, M., 1926, *Z. Phys.*, **40**, 421.
Bouwkamp, C. J., 1940, *Physica*, **7**, 485.
Breuninger, H. W., 1938, *Thesis*, University of Jena; 1939, *Ann. Phys., Lpz.*, **35**, 238.
Debye, P., 1909, *Ann. Phys., Lpz.*, **30**, 755.
Fabry, C., 1893, *J. Phys. Theor. Appl.* (3), **2**, 22.

---

† These follow from the Jacobi identities (Watson 1922, p. 22).

Fokker, A. D., 1923, *Physica*, **3,** 334; 1924, *Ibid.*, **4,** 166.
Gouy, L. G., 1890, *C. R. Acad. Sci., Paris*, **110,** 1251; 1891, *Ann. Chim. (Phys.)*, **24,** 145.
Ignatowsky, B. C., 1919, *Trans. State Optical Institute, Petrograd*, **1,** III.
Joubin, P., 1892, *C.R. Acad. Sci., Paris*, **115,** 932.
Julius, W. H., 1895, *Archives Néerl. Sci.*, **28,** 221, 226.
Linfoot, E. H., and Wolf, E., 1952, *Mon. Not. Roy. Astr. Soc.*, **112,** 452.
von Lommel, E., 1885, *Abh. Bayer Akad. Wiss.*, **53,** 233.
Maréchal, A., 1948, *Rev. Opt. (Theor. Instrum.)*, **27,** 73.
Picht, J., 1930, *Z. Phys.*, **65,** 14.
Pohl, R. W., 1948, *Optik*, 7th and 8th Ed. (Berlin : Springer), p. 353.
Reiche, F., 1909 a, *Ann. Phys., Lpz.*, **29,** 65; 1909 b, *Ibid.*, **29,** 401.
Rubinowicz, A., 1938, *Phys. Rev.*, **54,** 931.
Sagnac, G., 1903, *J. Phys. (Théor. Instrum.)*, **2,** 721; 1904 a, *C.R. Acad. Sci., Paris*, **138,** 479, 619, 678; 1904 b, *Boltzmann Festschrift*, 528.
Struve, H., 1886, *Mém. de l'Acad. de St. Petersb.* (7), **34,** No. 5, 1.
Toraldo di Francia, G., 1942, *Ottica*, **7,** N.2.
Watson, G. N., 1922, *Bessel Functions* (Cambridge : University Press).
Zeeman, P., 1897/98, *Versl. Gewone Vergad. Akad. Amst.*, **6,** 11; 1900, *Phys. Z.*, **1,** 542; 1901, *Archives Néerl. Sci.*, **4,** 314, 318.
Zernike, F., and Nijboer, B. R. N., 1949, Contribution to *Théorie des Images Optiques* (Paris : Éditions de la Revue d'Optique).

RESEARCH NOTES

Paramagnetic Resonance of Manganese in Single Crystals of Alkali Halides

By P. A. FORRESTER and E. E. SCHNEIDER
Department of Physics, King's College in the University of Durham, Newcastle-upon-Tyne

MS. received 23rd February 1956

§ 1. INTRODUCTION

We should like to report on the present state of researches on paramagnetic resonance in single crystals of alkali halides containing manganese. The experiments have been carried out at 290°K and at 80°K with a sensitive superheterodyne spectrometer (Schneider and England 1951) working at 9500 Mc/s. The account of the original work on the subject by Schneider and Caffyn (1955) appears to have escaped the attention of Oshima and others (1955) who have reported observations on the resonance in NaCl–Mn with an apparatus of low resolution which did not allow the detection of any fine structure in the resonance spectrum. Our work on crystals of NaCl and KCl grown from the melt and crystals of KCl, KBr and KI grown from solution have been supplemented by the results of Low on crystals of NaCl and KBr grown from solution which are reported in the following Note (Low 1956).

§ 2. EXPERIMENTAL RESULTS

(*a*) *NaCl–Mn*.

Crystals of NaCl grown from the melt containing 10^{-4} to 10^{-6} parts of Mn normally exhibit a single resonance peak of a width of about 400 gauss at half maximum and a *g*-factor close to 2·015. At concentrations of Mn larger than 10^{-4} we have observed widths ranging from 180 to 420 gauss.

The most significant observations concern the resonance pattern in NaCl–Mn after heat treatment of the crystals: a large number of lines, about 15 gauss wide, most of which depend on the orientation of the crystals in the magnetic field, appear superimposed on the broad peak. The line spectrum can be detected after heating to temperatures around 50°C, but in order to obtain a sufficient intensity (integrated intensity about equal to that of the broad line) a heating to 300–500°C and subsequent cooling to room temperature is required. The intensity of the line spectrum decays at room temperature within a period of several days leaving only the single broad peak.

Six strong lines can be identified in the spectrum which do not change with the orientation of the specimen in the field. They are interpreted on the basis of a spin-Hamiltonian $g\beta\mathbf{S}.\mathbf{B}+A\mathbf{S}.\mathbf{I}$ as the hyperfine components corresponding to the nuclear spin of Mn, $I=5/2$. The usual analysis of the line positions taking account of second-order terms in $A^2/g\beta$ leads to a hyperfine structure splitting constant $A=88{\cdot}7\pm0{\cdot}5$ gauss (as reported earlier, Schneider 1955) and an (isotropic) *g*-factor $g=2{\cdot}0011\pm0{\cdot}0005$. If the crystals are quenched from very high temperatures not far below the melting point, these six orientation independent hyperfine structure lines are predominant and have an integrated intensity which is at least 5% of that of the broad peak.

(*b*) *KCl–Mn grown from the melt.*

In crystals of KCl–Mn grown from the melt a single resonance peak is observed which is appreciably narrower than the single broad peak in NaCl–Mn. The line shape is typical of an exchange narrowed resonance and the width depends upon the thermal annealing history. The actual value varies between about 60 and 45 gauss at the half-height; after heating at about 600°c for several hours the line is always reduced to the minimum width of 45 ± 3 gauss at a g-value of $2{\cdot}002 \pm 0{\cdot}001$. These lines are permanent and show no sign of broadening with time, even after several months at room temperature. The form and position of the lines are independent of the concentration of Mn over the range 10^{-2} to 10^{-6}. The spectrum at 80°K is identical with that at 290°K.

In some crystals of KCl–Mn which had not been annealed sufficiently a broad line, about 200 gauss wide, has been observed. In such crystals very small red-coloured specks can be seen under a high power microscope. After annealing, the specks disappear and the broad resonance line is replaced by the characteristic narrow line. If the crystals are grown too rapidly the Mn^{2+} ions probably enter in microscopic inclusions and diffuse into the lattice only during a subsequent annealing.

Apart from these external effects thermal treatment of KCl–Mn crystals grown from the melt has no influence on the nature of the resonance spectrum. Even crystals quenched after prolonged heating near the melting point exhibit only the single resonance peak and no trace of a line spectrum.

(*c*) *KCl–Mn* (*KBr, KI*) *grown from solution.*

The spectrum of crystals of KCl–Mn grown from aqueous solution is entirely different from that of crystals grown from the melt. It consists of six hyperfine lines at $g = 2{\cdot}0047 \pm 0{\cdot}0005$, with a hyperfine structure splitting constant $A = 95{\cdot}0 \pm 0{\cdot}5$ gauss, which is somewhat larger than the value for NaCl–Mn. However, in contrast to the corresponding case of NaCl–Mn (Low 1956), there is no evidence of a single broad resonance. The crystals so far obtained, containing about 10^{-4} parts Mn, are small and the lines are about 50 gauss wide. With the resultant lack of resolution no further lines which might correspond to the orientation dependent lines of quenched NaCl–Mn have been observed. If crystals of KCl–Mn grown from solution are annealed for several hours at 500°c the hyperfine lines disappear and are replaced by the narrow line characteristic of crystals grown from the melt. This conversion is irreversible: it is impossible to re-establish the conditions responsible for the hyperfine structure.

The spectra of Mn ions in KBr and in KI grown from solution are similar to that in KCl in agreement with the results of Low (1956).

We have made the remarkable observations that in all three potassium halides the hyperfine structure lines disappear upon cooling the crystals to 80°K and reappears on warming to room temperature.

§ 3. Discussion

The three types of paramegnetic resonance spectra of Mn appearing alone or in combination, namely the single peak resonance, the simple orientation independent hyperfine structure spectrum, and the complex orientation dependent line spectrum, can be associated with three different types of habitat of the foreign divalent ion in the monovalent host lattice.

Two alternative models have been discussed by Schneider and Caffyn to account for the broad resonance peak in NaCl–Mn. In one of them the Mn^{2+} ions are assumed to be in random positions at or near internal boundaries, made up of dislocations or other imperfections, so that the consequent non-cubic fields, of random orientation and strength, lead by fine structure interaction to a broadening of the resonance and hence to a smoothing out of the hyperfine structure. On the other hand, Mn^{2+} ions could aggregate at or near internal boundaries and the hyperfine structure be suppressed by exchange effects. In view of the narrower resonances which have now been observed in NaCl, and in particular of the results on KCl, we would consider now that aggregation and exchange effects must be responsible, at least partially, for the main resonance both in crystals grown from the melt and from solution. The experiments on KCl–Mn show that during the process of growth from the melt and during annealing at high temperature of crystals grown from solution, the Mn^{2+} ions diffuse within the crystals until they are trapped at internal boundaries, where they cluster into aggregates: the shape of the resonance line, showing absorption far out into the wings, is evidence for a strong exchange interaction which suppresses all structure and narrows the line to a value well below that expected from the spin–spin interaction.

The simple orientation independent hyperfine structure of six lines must be ascribed to isolated Mn ions in cation sites of the lattice where the crystalline field is essentially cubic. At first sight it appears surprising that the pure hyperfine structure spectrum is exhibited strongly in the potassium halides, since on the basis of the criterion for the incorporation of foreign ions into ionic lattices deduced by Haven (1950) one could not expect the small Mn^{2+} ion, $r = 0{\cdot}80$ Å, to replace the larger K^+, $r = 1{\cdot}33$ Å, in perfect regions of the lattice. The misfit certainly explains why at high temperatures Mn ions have a greater tendency to aggregate irreversibly at internal boundaries in KCl than in NaCl which is known to take up Mn in solid solution. However, one must not overlook that the disparity between the sizes of foreign and host ion causes the mobility of Mn ions at or near room temperature to be very much lower in the potassium halides than in the sodium halides, a fact which has also been deduced by Seitz (1954) from optical investigations. Thus, in the process of growth of the potassium halides from solution, Mn ions may be deposited in the distorted regions within a few lattice distances from internal boundaries without being able to diffuse to the boundaries and to form aggregates. In such positions the Mn ions would be subjected to sufficiently strong random axial fields, superimposed on the average cubic field produced by the surrounding lattice ions, to account for the large width and gaussian shape of the observed six hyperfine structure lines. In this case the resonance spectrum must be associated with the main series of transition, $m_S = +1/2 \longleftrightarrow -1/2$ which is affected only in second order by axial fields. The other four series of transitions, $m_S = \pm 5/2 \longleftrightarrow \pm 3/2$, $\pm 3/2 \longleftrightarrow \pm 1/2$ would be broadened so much by first-order fine structure interaction with the axial fields to be unobservable.

One could go even further and explain the disappearance of the resonance at low temperatures as a broadening by very strong random fields effects. One would then have to invoke a kind of motional narrowing by lattice vibrations by which the effect of the random fields is reduced at normal temperatures.

The hyperfine structure spectrum of Mn in crystals of NaCl grown from solution which is described in the following Note (Low 1956) represents the other extreme. The small intensity of the lines compared with that of the broad peak shows that the majority of the Mn ions diffuse to internal boundaries and their small width demonstrates that the remainder, about 0·02% of the total number of Mn ions, are in a closely perfect crystalline environment.

The heat-treated NaCl–Mn crystals grown from the melt provide an intermediate case. The larger residual width of the orientation independent hyperfine structure lines suggests that the Mn ions are still close enough to the internal boundaries, from which they have been dissociated thermally, for a broadening effect by random axial fields to occur. Their integrated intensity after normal heat treatment shows that the number of isolated Mn ions is about 1% of the total concentration in agreement with the corresponding number of unattached additional positive ion vacancies deduced from ionic conductivity measurements (Schneider and Caffyn 1955, Parker 1954). In crystals quenched from very high temperatures the fraction of isolated Mn ions is about five times larger.

It is interesting to consider the values of the hyperfine structure splitting constant A in the light of our interpretation of the spectrum. According to van Wieringen (1955) the configurational interaction responsible for the hyperfine structure of Mn^{2+} decreases with increasing covalent bonding of the ion. The A values for all alkali halides lie between that for ZnS–Mn, $A = 68{\cdot}1$ gauss (Schneider and England 1951) and that for ZnF_2–Mn, $A = 104$ gauss (Tinkham 1955), where a secondary hyperfine structure due to covalent interaction with the F-nuclei occurs. This points to a substantial covalent contribution to the bonding of the Mn^{2+} ion in the alkali halide host lattice whose variation from crystal to crystal fits very satisfactorily into the general picture presented here: the A value is largest, the covalent contribution smallest, in the potassium halides where the Mn ion enters only with difficulty into the lattice, while A is smaller, the covalent contribution larger, in NaCl where the Mn ion participates more intimately in the normal bonding with adjacent lattice ions. In this way, even the small difference between our values and Low's for the hyperfine structure splitting of Mn in NaCl becomes significant since it reflects the more nearly perfect substitutional incorporation of Mn ions into the lattice of crystals grown from solution.

Work is still in progress on the analysis of the third type of spectrum, the orientation dependent line spectrum of NaCl–Mn which we believe to be the most direct evidence obtained so far for the existence of complexes of divalent impurity ions, in the present case Mn^{2+} ions, and adjacent positive ion vacancies. In such complexes, an axial field produces a fine structure which is large compared with the hyperfine structure, in fact so large ($D \sim 500$ gauss) that because of second order effects even the lines due to the transitions $m_s = -\frac{1}{2} \longleftrightarrow m_s = +\frac{1}{2}$ depend on orientation.

REFERENCES

HAVEN, Y., 1950, *Rec. Trav. Chim. Pays-Bas*, **69**/2, 1259.

LOW, W., 1956, *Proc. Phys. Soc.* B, **69**, 837.

OSHIMA, K., ABE, H., NAGANO, H., and NAGUSA, K., 1955, *J. Chem. Phys.*, **23**, 1721.

PARKER, N. S., 1954, *Ph.D. Thesis*, University of Durham.

SCHNEIDER, E. E., 1955, *Disc. Faraday Soc.*, **19** (Microwave and Radio-frequency Spectroscopy), p. 174.

SCHNEIDER, E. E., and CAFFYN, J. E., 1955, *Defects in Crystalline Solids*, Report of 1954 Bristol Conference (London: Physical Society), p. 74.
SCHNEIDER, E. E., and ENGLAND, T. S., 1951, *Physica*, **17**, 221.
SEITZ, F., 1954, *Rev. Mod. Phys.*, **26**, 7.
TINKHAM, M., 1955, *Disc. Faraday Soc.*, **19** (Microwave and Radiofrequency Spectroscopy), p. 174.
VAN WIERINGEN, J. S., 1955, *Disc. Faraday Soc.*, **19** (Microwave and Radiofrequency Spectroscopy), p. 118.

Paramagnetic Resonance Spectrum of Manganese in Single Crystals of Alkali Halides Grown from Solutions†

BY W. LOW‡
The Enrico Fermi Institute for Nuclear Studies, University of Chicago, Ill.

Communicated by E. E. Schneider; MS. received 23rd *February* 1956

THE paramagnetic resonance spectrum of manganese has been observed at 3 cm and 1 cm in single crystals of NaCl and KBr grown from aqueous solutions and containing 0·001–0·02% Mn by weight.

The spectrum in NaCl can be divided into two parts: (*a*) a strong line of width at half intensity of about 220 ± 20 gauss, approximately Lorentzian shape and $g=2{\cdot}010\pm0{\cdot}005$, in agreement with the results of Oshima *et al.* (1955); (*b*) superimposed on this line is a hyperfine structure of six lines due to the known spin of 5/2 of ^{55}Mn. The width of the hyperfine lines at half intensity is 7–8 gauss. The spectrum of the hyperfine lines is described by the formula

$$h\nu=g\beta H+Am+\tfrac{1}{2}\,\frac{A^2}{h\nu}\,\{I(I+1)-m^2\}$$

with $g=2{\cdot}0022\pm0{\cdot}0008$, $A=86{\cdot}6\pm0{\cdot}2$ gauss. No fine structure was observed.

The spectrum in KBr (as well as KCl) consists of six hyperfine lines. The hyperfine splitting constant is $A=94{\cdot}95\pm0{\cdot}3$ gauss, and at $g=2{\cdot}0043\pm0{\cdot}0005$. The width of the lines at half intensity is about 40–45 gauss. The results on KBr are very similar to those on KCl grown from solutions as reported by Forrester and Schneider (1956), and also observed by the author.

The spectra in NaCl and KBr grown from solutions differ, therefore, in four aspects: (i) in NaCl there is an additional intense and wide line; (ii) in NaCl the hyperfine lines are considerably narrower; (iii) the g factor in KCl and KBr is significantly larger than in NaCl, and larger than g_e, that of the free electron; (iv) the hyperfine splittings of NaCl and KBr differ by about 8%.

The significance of the intense and broad line in NaCl has been discussed by Schneider and Caffyn (1955). They find that the wide and strong line is dependent on the concentration of divalent impurities, the large the concentration the narrower the line. In crystals grown from solutions the line width is approximately independent of the concentration. This seems to point, therefore, to significant differences in the incorporation of divalent impurities, depending on the method by which the crystals are grown.

† Supported by the United States Atomic Energy Commission.

‡ On leave of absence from the Hebrew University, Jerusalem.

The hyperfine structure constant of 86·6 and the g factor of 2·0022 in NaCl are very close to those found in single crystals of MgO : Mn (concentration 0·001%, line width 3–4 gauss, $A=86{\cdot}8$, $g=2{\cdot}0014\pm0{\cdot}0005$) (Low 1956). The g factors in KBr and KCl are very much larger than in nearly all Mn salts investigated. The deviations of the g factor from the precise value in MgO can be attributed to both the crystalline field and the spin–orbit interaction. In all other manganese salts of various crystalline symmetries these two effects have made g less than g_e. While it may be difficult to decide which of these two factors is the main cause for the increase in the g factor, it seems likely that in the case of KBr and KCl a change in the spin–orbit coupling may be the main contributor. This may possibly be due to association of the Mn ion with adjacent electrons.

The wide line in NaCl was about a hundred times more intense than the hyperfine components which made the observations of the latter rather difficult. A technique developed by Hutchison and Noble (Noble 1955) was used in detecting the hyperfine lines. The magnetic field is modulated over a small range and the line is detected at the second harmonic of the modulation frequency. This discriminates against any slow changes in the slope of the wide line but detects the relatively narrow peaks of the hyperfine structure. The second derivative of the absorption line is recorded.

ACKNOWLEDGMENT

The author would like to thank Professor C. A. Hutchison, Jr., for helpful discussion and interest.

REFERENCES

FORRESTER, P. A., and SCHNEIDER, E. E., 1956, *Proc. Phys. Soc.* B, **69, 833.**

LOW, W., 1956, *Phys. Rev.*, **101,** 1827.

NOBLE, G., 1955, *Thesis*, University of Chicago.

OSHIMA, K., ABE, H., NAGANO, H., and NAGUSA, K., 1955, *J. Chem. Phys.*, **23,** 1721.

SCHNEIDER, E. E., and CAFFYN, J. E., 1955, *Defects in Crystalline Solids*, Report of 1954 Bristol Conference (London : Physical Society), p. 74.

Equation of State of Solids

BY R. S. SHARMA[†]
Physics Department, Royal Holloway College, London

Communicated by S. Tolansky; MS. received 22*nd December* 1955,
and in final form 12*th March* 1956

In a recent communication (Dayal and Sharma 1955) it has been shown that it would be advantageous to base the equation of state of solids on the experimental data of thermal expansion and specific heat, thus avoiding the necessity of assumptions regarding any theory of lattice vibrations or the potential expressions. The total pressure of a solid can be considered to be made up of two parts p_1 and p_2, of which p_1 can be interpreted as a pressure due to the non-vibrating lattice at $T=0°\text{K}$, while p_2 results from the thermal vibrations. For all volumes the sum of these two is equal to the outside pressure, which is almost equal to zero in the scale of pressures considered. Thus for all volumes measured at atmospheric pressure, the two are equal and opposite to each other. The static pressures $\bar{\gamma}E/V$ of eight solids determined from their thermal pressures have been compared with the derivatives of the potential energy as derived from the interatomic potential used by Fürth (1945) and close agreement between them revealed. Similar calculations are now reported for the solids, lithium and sodium, which belong to the body-centred cubic class.

[†] Permanent address, Department of Physics, Banaras Hindu University, India.

Lithium: The accurate measurements of the specific heats of Li in range (15–300°K) are due to Simon and Swain (1935). Although the coefficients of thermal expansion were measured by Simon and Bergmann (1930), these have been taken from Lord (1941) who interpolated the Simon and Bergmann data for that purpose. E (thermal energy) is calculated from its vibration frequencies as given by Dayal (1944). In order to obtain the average value of the Grüneisen constant γ, k_0 the compressibility at absolute zero is calculated from the elastic constants obtained theoretically by Füchs (1936). $(m+n+3)/6$ is taken equal to 1·83, as given by Grüneisen (1926), for calculating the values of γ for the various volumes.

Sodium: The data of Simon and Zeidler (1926) on the specific heat, and of Quimby and Seigel (1938) on the thermal expansion have been used for the present calculations. For the value of the specific heat C_V at the lowest temperature (80°K) the Debye characteristic temperature is calculated and this in turn is used to obtain the thermal energy at that temperature. The values of thermal energies at higher temperature are obtained from the *Landolt–Börnstein Tables*; k_0 at absolute zero is given by B nder (1939); $(m+n+3)/6$ is assumed equal to 1·92 as quoted by Grüneisen (1926).

From the table, it appears that the agreement is excellent in the case of lithium, but for sodium the two sets of values do not compare well at low temperatures.

Substance	T (°K)	V (cm³)	E (cal. (g atom)$^{-1}$)	$\bar{\gamma}$	$-p_1$ (10^8 dyn cm^{-2})	
Lithium					Fürth	$\bar{\gamma}E/V$
($\Omega=36·0$ kcal (g atom)$^{-1}$,	160	12·884	343	0·885	11	10
$V_0=12·728$ cm³, $m=1·2$,	200	12·940	525	0·957	16	15
$n=6$)	240	13·003	725	0·988	22	23
	300	13·097	1034	1·11	30	33
Sodium						
($\Omega=26·2$ kcal (g atom)$^{-1}$,	95	22·743	344	1·25	8	5
$V_0=22·61$ cm³, $m=2·5$,	110	22·773	422	1·279	10	6
$n=6$)	130	22·842	538	1·524	13	8
	200	23·181	958	1·366	19	20
	300	23·70	1547	1·341	36	39

Ω = sublimation energy.

This discrepancy can be attributed to the uncertainty in the measurements of the thermal expansions at the temperatures near the absolute zero. A slight error in the value of thermal expansions causes a slight variation in the volume but involves a considerable change in the pressures.

My sincere thanks are due to Dr. B. Dayal of Banaras Hindu University (India) for helpful discussions.

REFERENCES

BENDER, O., 1939, *Ann. Phys., Lpz.*, **34**, 339.
DAYAL, B., and SHARMA, R. S., 1955, *Proc. Phys. Soc.* B, **68**, 1049.
DAYAL, B., 1944, *Proc. Indian Acad. Sci.* A, **20**, 24.
FÜCHS, K., 1936, *Proc. Roy. Soc.* A, **157**, 444.
FÜRTH, R., 1945, *Proc. Roy. Soc.* A, **183**, 87.
GRÜNEISEN, E., 1926, *Handbuch d. Physik*, Vol. X.
LORD, R. C., Jr., 1941, *J. Chem. Phys.*, **9**, 693.
QUIMBY, S. L., and SEIGEL, K. S., 1938, *Phys. Rev.*, **54**, 383.
SIMON, F., and BERGMANN, M., 1930, *Z. Phys. Chem.* B, **8**, 255.
SIMON, F., and SWAIN, R. C., 1935, *Z. Phys. Chem.* B, **28**, 189.
SIMON, F., and ZEIDLER, N., 1926, *Z. Phys. Chem.* B, **123**, 383.

On the Crystal Structure of the Rare Gases

BY R. U. AYRES AND R. H. TREDGOLD†
University of Maryland

MS. received 9*th March* 1956 *and in revised form* 22*nd March* 1956

§ 1. INTRODUCTION

IT is well known that, with the exception of helium, the rare gases crystallize in the face-centred cubic structure. It is however difficult to understand why this structure, as opposed to the close-packed hexagonal structure, should be preferred. Investigations of various possible two-body force laws have been carried out (for a summary see Dawson and Jansen 1955) and it has been shown that for any reasonable form of these laws there is only a negligible difference in energy between the two structures. Axilrod (1951 a, b) investigated the effect of the three-body forces arising from the interaction of the dipole moments of sets of three atoms. These forces appear when the calculation of Van der Waals forces is carried through to third order perturbation. He found that, whereas these forces favour the face-centred cubic structure, their effect is too small (of the order 10^{-4} the cohesive energy) to be important. Barron and Domb (1955) examined the influence of the zero point energy of the lattice vibrations on the relative stabilities of the two structures but found that the energy difference so obtained was negligible.

The other possible effects which are likely to influence the lattice energy and which have so far not been examined are: (1) the interaction of the repulsive first order overlap energy with the triple dipole forces, (2) higher order multipole three-body forces, (3) four-body and higher order many-body forces, (4) the dependence of the repulsive forces on crystal symmetry. The effect of (3) and (4) cannot be ruled out entirely but qualitative considerations indicate that it is unlikely that they will play a decisive role.

We have, therefore, made a quantitative investigation of effects (1) and (2). It is not intended to give a detailed description of the study of the effect of overlap on three-body forces as it was found that, though the resulting energy term favoured the cubic structure, it was of the same order of magnitude as the effect studied by Axilrod (1951 a, b) and therefore not likely to be important.

---

† Now at the Department of Physics, University of Sheffield.

§ 2. HIGHER ORDER THREE BODY FORCES

The higher order triple multipole forces are in general unimportant owing to their extremely short range. However, the dipole–dipole–quadrupole forces are not negligible and, owing to their marked angular dependence, the resulting energy terms arising from them in the two close-packed structures differ appreciably as will be seen below.

We consider three atoms located at lattice sites 1, 2 and 3 and write the interaction Hamiltonian in the form

$$H' = H_{\mathrm{dd}}(12) + H_{\mathrm{dq}}(23) + H_{\mathrm{dq}}(31) \qquad \ldots\ldots(1)$$

indicating dipoles at 1 and 2 and a quadrupole at 3.

$$H_{\mathrm{dd}}(12) = \frac{e^2}{R_{12}{}^3}\left[\mathbf{r}_1 \cdot \mathbf{r}_2 - 3\,\frac{(\mathbf{r}_1 \cdot \mathbf{R}_{12})(\mathbf{r}_2 \cdot \mathbf{R}_{12})}{R_{12}{}^2}\right]$$

and

$$\begin{aligned} H_{\mathrm{d_q}}(12) = \frac{3e^2}{2R_{12}{}^4}\Bigg\{ & \frac{[\mathbf{r}_1{}^2\mathbf{r}_2 + 2(\mathbf{r}_1 \cdot \mathbf{r}_2)(\mathbf{r}_1 - \mathbf{r}_2) - \mathbf{r}_2{}^2\mathbf{r}_1] \cdot \mathbf{R}_{12}}{R_{12}} \\ & - 5\,\frac{(\mathbf{r}_1 \cdot \mathbf{R}_{12})(\mathbf{r}_2 \cdot \mathbf{R}_{12})(\mathbf{r}_1 - \mathbf{r}_2) \cdot \mathbf{R}_{12}}{R_{12}{}^3}\Bigg). \end{aligned} \qquad \ldots\ldots(2)$$

The third-order perturbation energy is

$$E_3 = \sum_{j,k}{}'' \frac{H_{0j}{}' H_{jk}{}' H_{k0}{}'}{(E_0 - E_j)(E_0 - E_k)} \qquad \ldots\ldots(3)$$

the other terms which usually appear in third-order perturbation being zero in this case.

Making the approximation

$$E_j - E_0 = E_{\mathrm{ionization}} \qquad \ldots\ldots(4)$$

and using the fact that

$$\sum_{j,k}{}'' H_{0j}{}' H_{jk}{}' H_{k0}{}' = \langle H'^3 \rangle_{00} \qquad \ldots\ldots(5)$$

we find from equation (3)

$$E_3 = \frac{\langle H'^3 \rangle_{00}}{E_{\mathrm{ion}}{}^2}. \qquad \ldots\ldots(6)$$

Inserting (1) in (6) we can show that the only non-zero contribution arises from the cross term $\langle H_{\mathrm{dd}}(12) H_{\mathrm{dq}}(23) H_{\mathrm{dq}}(31) \rangle$ for all permutations 1, 2, 3.

Denoting the distances between the atoms by R_{12}, R_{23}, R_{31} and the interior angles of the triangle on which they are situated by θ_1, θ_2, θ_3 we have

$$\begin{aligned} \langle H_{\mathrm{dd}}(12) H_{\mathrm{dq}}(23) H_{\mathrm{dq}}(31) \rangle = \frac{9}{4}\,\frac{e^2\alpha^2}{R_{12}{}^3 R_{23}{}^4 R_{31}{}^4} & \{(10\alpha^2 - \beta^2)(\cos\theta_3 + 3\cos\theta_1\cos\theta_2) \\ & - 25\beta^2(\cos\theta_3 + 3\cos\theta_1\cos\theta_2)\cos^2\theta_3 \\ & - 50(\beta^2 - 2\alpha^2)(\cos\theta_3 + 3\cos\theta_1\cos\theta_2)\sin\theta_1\sin\theta_2\cos\theta_1\cos\theta_2 \\ & + 10(\beta^2 + 3\alpha^2)(\sin^2\theta_1 + \sin^2\theta_2)\sin\theta_1\sin\theta_2 \\ & + 20\beta^2(\cos^2\theta_1 + \cos^2\theta_2)\cos\theta_1\cos\theta_2 - 60\alpha^2\sin\theta_1\sin\theta_2\} \end{aligned} \qquad \ldots\ldots(7)$$

where

$$\alpha = \langle x^2 \rangle = \langle y^2 \rangle = \langle z^2 \rangle$$

$$\beta^2 = \langle x^4 \rangle = \langle y^4 \rangle = \langle z^4 \rangle.$$

The appropriate lattice sums were carried out for the face-centred cubic and close-packed hexagonal lattices on a high speed computing machine. The convergence is rapid so that a radius of two nearest neighbour distances from the central atom is a sufficient limit for the sums. α and β^2 can be calculated to sufficient accuracy by using the outer shell wave functions. For argon one thus finds

$$\alpha = 4{\cdot}801 a_0^2, \qquad \beta^2 = 62{\cdot}001 a_0^4 \qquad \ldots\ldots(8)$$

where a_0 is the Bohr radius. Using the accepted values for argon $r_0 = 7{\cdot}27 a_0$ and $E = 15{\cdot}68$ ev one obtains for the dipole–dipole–quadrupole energy

$$E_{\mathrm{cph}} - E_{\mathrm{fcc}} = -0{\cdot}000135 \text{ ev atom}^{-1} \qquad \ldots\ldots(9)$$

which is about 0·3% of the cohesive energy and *favours the hexagonal lattice.*

§ 3. Discussion

The above results indicate that either there is some other effect involved which has so far not been appreciated, or, possibly, that the hexagonal structure does represent the lowest energy state in solid rare gases but that at higher temperatures the face-centred cubic structure is stable and 'frozen in' on cooling. There is also a possibility that higher order triple multipole forces may be more important than qualitative considerations lead one to believe.

Acknowledgments

This research was supported by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

References

Axilrod, B. M., 1951 a, *J. Chem. Phys.*, **19**, 719; 1951 b, *Ibid.*, **19**, 724.
Barron, T. H. K., and Domb, C., 1955, *Proc. Roy. Soc.* A, **227**, 447.
Dawson, L. M., and Jansen, J., 1955, *J. Chem. Phys.*, **23**, 482.

The Influence of Interelectronic Collisions and of Surfaces on Electronic Conductivity

By H. FRÖHLICH, B. V. PARANJAPE, C. G. KUPER
and S. NAKAJIMA
Department of Theoretical Physics, University of Liverpool

MS. received 27th February 1956

§ 1.

In an electric conductor with finite cross section it must be possible to impose a surface condition on the current density **j**. Ohm's law does not permit this and cannot, therefore, be quite correct. It will be shown below that it has to be supplemented by terms describing a viscous flow.

Collisions between free electrons conserve momentum and do not, therefore, contribute to the ohmic resistivity. They do, however, make a contribution to the viscosity in a similar way to collisions between molecules in the internal friction of gases. The macroscopic equation for **j**, which will be derived below, should be able to account for surface effects of various kinds (effective conductivity of thin layers, anomalous skin effect, etc.) and also put interelectronic collisions into evidence whenever **j** depends on spatial coordinates. In metals, in view of the strong screening, such collisions are probably unimportant. In semiconductors, however, a density range exists where these collisions may be of considerable importance, as has been shown recently (Fröhlich and Paranjape 1956, § 1.2).

§ 2.

Consider the stationary case only, and assume a specimen bounded by the plane $z=0$, with an external field F_x in the x direction. The electronic distribution function f may depend on the space coordinate z as well as on the velocity coordinates $\mathbf{v}$ and is assumed to be normalized, such that

$$\rho(z)=\int f d^3\mathbf{v} \qquad \ldots\ldots(1)$$

is the number density of free electrons assumed independent of x and y; f satisfies the transport equation

$$\frac{eF_x}{m}\frac{\partial f}{\partial v_x}+\frac{eF_z}{m}\frac{\partial f}{\partial v_z}+v_z\frac{\partial f}{\partial z}=\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} \qquad \ldots\ldots(2)$$

where the right-hand side represents the rate of change of f due to collisions of all kinds. They will be treated in a phenomenological way by assuming the existence of relaxation times.

The components of the current density are

$$j_x(z)=e\int v_x f d^3\mathbf{v}\,; \quad j_z=e\int v_z f d^3\mathbf{v}\equiv 0\,; \quad j_y\equiv 0. \qquad \ldots\ldots(3)$$

j_z must be zero, but a field F_z in the z direction, in general, must be contemplated in order to compensate a diffusion current if the density ρ is z-dependent.

The distribution function f can be developed in a series of spherical harmonics (in velocity space) whose coefficients may be z-dependent. Apart from a zero order term, this series must have a single first order term, proportional to v_x (so that $j_z\equiv j_y\equiv 0$). The possibility of spatial diffusion in the z direction then requires two second-order terms with v_xv_z and $v_x^2-v^2/3$ symmetry respectively. The series is assumed to converge quickly so that third or higher order harmonics can be neglected. This assumption represents the main limitation of the present method.

Interelectronic collisions do not alter the total momentum density $\int v_x f d^3\mathbf{v}$. Hence the average $\bar{\tau}_c$ of a relaxation time τ_c responsible for ordinary conduction can be defined by

$$e\int v_x\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} d^3\mathbf{v}=-\frac{j_x(z)}{\bar{\tau}_c}, \qquad \ldots\ldots(4)$$

where τ_c arises from collision with lattice displacements and defects. While thus the v_x term of f is not affected by interelectronic collision, this does not hold for other terms, such as the v_xv_z term. Thus if τ_{el} is a relaxation time describing interelectronic collisions and if

$$1/\tau=1/\tau_c+1/\tau_{el}, \qquad \ldots\ldots(5)$$

then an appropriate average $\bar{\tau}$ of τ can be defined by

$$\int v_xv_z\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} d^3\mathbf{v}=-\frac{1}{\bar{\tau}}\int v_xv_z f d^3\mathbf{v}. \qquad \ldots\ldots(6)$$

Now multiply (2) by v_x and v_xv_z respectively and integrate over $d^3\mathbf{v}$. Then

$$-\frac{e^2F_x}{m}\rho(z)+e\frac{\partial}{\partial z}\int v_xv_z f d^3\mathbf{v}=-\frac{j_x(z)}{\bar{\tau}_c} \qquad \ldots\ldots(7)$$

and

$$-\frac{e^2F_z}{m}j_x(z)+e\frac{\partial}{\partial z}\int v_xv_z^2 f d^3\mathbf{v}=-\frac{e}{\bar{\tau}}\int v_xv_z f d^3\mathbf{v}. \qquad \ldots\ldots(8)$$

Here $v_x v_z^2$ can be split into a term with v_x-symmetry and a negligible third-order term so that

$$e\int v_x v_z^2 f d^3\mathbf{v} = \alpha\overline{v^2} j_x(z) \qquad \ldots\ldots(9)$$

where α is a numerical factor (whose exact value depends on the velocity dependence of τ_c), and $\overline{v^2}$ is defined by

$$\overline{v^2}\rho(z) = \int v^2 f d^3\mathbf{v}. \qquad \ldots\ldots(10)$$

With (7)–(9) thus

$$j_x - \frac{e^2\bar{\tau}_c\rho}{m}F_x - \alpha\overline{v^2}\bar{\tau}_c\frac{\partial}{\partial z}\left(\bar{\tau}\frac{\partial j_x}{\partial z}\right) + \frac{e^2\bar{\tau}_c}{m}\frac{\partial}{\partial z}(\bar{\tau} j_x F_z) = 0 \qquad \ldots\ldots(11)$$

where j_x, F_z, $\bar{\tau}_c$ and $\bar{\tau}$ may all depend on ρ and hence on z. F_z can be expressed in terms of ρ, for multiplication of (2) by v_z and integration leads to

$$\frac{eF_z(z)}{m}\rho(z) = \frac{\overline{v^2}}{3}\frac{\partial\rho}{\partial z}. \qquad \ldots\ldots(12)$$

The z-dependence of ρ has to be obtained in an independent way from conditions imposed on the substance (e.g. photoconductivity).

§3.

In a phenomenological description one will want to introduce a conductivity σ and a viscosity η by

$$\sigma = e^2\bar{\tau}_c\rho/m; \qquad \eta = \tfrac{1}{2}m\overline{v^2}\rho\bar{\tau}. \qquad \ldots\ldots(13)$$

Generalization of (11) and (12) to arbitrary geometrical conditions and elimination of the internal field F_z then leads to

$$\mathbf{j} - \sigma\mathbf{F} + \frac{2}{e^2}\frac{\sigma}{\rho}\left\{\alpha\,\mathrm{curl}\left(\frac{\eta}{\rho}\,\mathrm{curl}\,\mathbf{j}\right) + \frac{1}{3}\,\mathrm{curl}\,\frac{\eta}{\rho}\,[\mathbf{j}\times\mathrm{grad}\log\rho]\right\} = 0. \qquad \ldots\ldots(14)$$

In conditions in which ρ is a constant independent of the spatial coordinates, σ and η are also constants and (14) reduces to

$$\mathbf{j} - \sigma\mathbf{F} + \lambda^2\,\mathrm{curl}\,\mathrm{curl}\,\mathbf{j} = 0. \qquad \ldots\ldots(15)$$

Here

$$\lambda^2 = \alpha\bar{l}_c\bar{l}, \quad \bar{l}_c = \bar{\tau}_c(\overline{v^2})^{1/2} \quad \text{and} \quad \bar{l} = \bar{\tau}(\overline{v^2})^{1/2} \qquad \ldots\ldots(16)$$

are appropriately averaged mean free paths. It should be noticed that equation (15) differs from Ohm's law by the term characteristic for viscous flow even in the absence of interelectronic collisions, i.e. when $\bar{l} = \bar{l}_c$. This permits a simple derivation of the influence of surface on the effective conductivity of thin layers which depends, of course, on the surface condition imposed on $\mathbf{j}$. If, for instance, it is assumed that electrons are scattered diffusely by the surface, then the surface condition is

$$\mathbf{j}(\mathbf{s}) = -\lambda\frac{\partial\mathbf{j}(\mathbf{s})}{\partial n} \qquad \ldots\ldots(17)$$

where the coordinates $\mathbf{s}$ define the surface and $\partial/\partial n$ represents differentiation perpendicular to it in the outward direction. For a wire of radius d, $0 \leqslant r \leqslant d$, (15) and (17) are solved by

$$i_x = \sigma F\left(1 - \frac{J_0(ir/\lambda)}{J_0(id/\lambda) - iJ_1(id/\lambda)}\right) \qquad \ldots\ldots(18)$$

where the J's are Bessel functions. For the effective conductivity $\bar{\sigma}$ one then obtains

$$\frac{\bar{\sigma}}{\sigma}=\frac{2}{\sigma F d^2}\int_0^d j_z r\,dr=1+\frac{2i\lambda}{d}\,\frac{\mathrm{J}_1(id/\lambda)}{\mathrm{J}_0(id/\lambda)-i\mathrm{J}_1(id/\lambda)}. \qquad \ldots\ldots(19)$$

In the limits $d \ll \lambda$ and $d \gg \lambda$ this expression becomes $\frac{1}{2}d/\lambda$ and $1-\lambda/d$ respectively and thus agrees (apart from numerical factors) with values obtained by Dingle (1950). For a slab of thickness d, on the other hand, $\bar{\sigma}/\sigma$ agrees only for $d>\lambda$ with the result by Fuchs (1938). For $d \ll \lambda$ it differs by a factor $\log(\lambda/d)$. In this case the distribution function f is so strongly asymmetrical (more than in the case of the wire) that it is no longer permissible to neglect third order and higher harmonics, as is done in the present paper.

One of the main applications of (14) should lead to measurement of the interelectronic collision probability in semiconductors. For it should be possible to devise a series of measurements such that ρ, σ and η are found experimentally. Hence one obtains $(\bar{\tau}_c/\bar{\tau})-1$, which is a measure for the relative importance of interelectronic collisions. Viscous effects would be particularly strong in photoconduction, where ρ should depend on the spatial coordinates in a similar way to the intensity of the ionizing radiation.

REFERENCES

DINGLE, R. B., 1950, *Proc. Roy. Soc.* A, **201,** 545.
FRÖHLICH, H., and PARANJAPE, B. V., 1956, *Proc. Phys. Soc.* B, **69,** 21.
FUCHS, K., 1938, *Proc. Camb. Phil. Soc.*, **34,** 100.

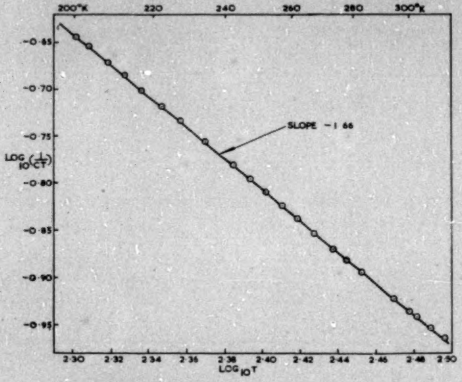
The Temperature Dependence of the Mobility of Electrons in Germanium

BY D. M. EVANS
Radio Research Station, Slough, Bucks.

Communicated by R. L. Smith-Rose; MS. received 9th January 1956

THE temperature dependence of the lattice-scattering mobility of electrons in germanium has been reported as being different from $T^{-1\cdot5}$ as predicted by the deformation-potential theory (Bardeen and Shockley 1950) and other theoretical derivations. The exponent of T was determined in the drift mobility experiments of Prince (1953) to be about $-1\cdot6$ while the more accurate conductivity measurements of Morin (1954) showed the exponent to be $-1\cdot66$. An exponent of $-2\cdot0$ has been reported (Many 1954) from drift mobility measurements, which differs markedly from the value determined by Prince and Morin. The theoretical consideration of the 'many-valley' model by Herring (1955) shows that the negative exponent can be increased to a value greater than the 1·5 of the simple theory.

The results of a determination of the temperature dependence of the mobility of electrons in p-type germanium by measurements on an n–p–n transistor are given here. Precisely the same method was used as has been previously described for the determination of the temperature dependence of the mobility of holes in n-type germanium by measurements on p–n–p transistors (Evans 1956). This method is dependent on the use of the equation for the current amplification factor given by the Shockley theory of junction transistors on the assumption of unit injection efficiency. An n–p–n alloy-type transistor for which the above theory was determined to be applicable over the temperature range 200 to 314°K was used. The results obtained are shown in the figure, in which the electron mobility μ_n is proportional to $1/CT$ where C is a variable arm of the bridge used to determine the current amplification factor and T the absolute temperature. The slope of the plot of $\log(1/CT)$ against $\log T$ gives the exponent of T, x (say), in the relation $\mu_n \propto T^{-x}$. The experimental points fit a straight line with a slope between $-1{\cdot}65$ and $-1{\cdot}67$; the best fit is obtained with a line of slope $-1{\cdot}66$. For the temperature range quoted above the mobility is considered to be entirely governed by lattice scattering. The results therefore illustrate that the temperature dependence of the lattice-scattering mobility of electrons in germanium is as $T^{-1\cdot 66}$ in agreement with the results of Morin.



Variation of electron mobility with temperature; $\mu_n \propto 1/CT$.

ACKNOWLEDGMENTS

The work described above was carried out as part of the programme of the Radio Research Board and this note is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

REFERENCES

BARDEEN, J., and SHOCKLEY, W., 1950, *Phys. Rev.*, **80,** 72.
EVANS, D. M., 1956, *J. Electronics*, **1,** 461.
HERRING, C., 1955, *Bell Syst. Tech. J.*, **34,** 237.
MANY, A., 1954, *Physica*, **20,** 985.
MORIN, F. J., 1954, *Phys. Rev.*, **93,** 62.
PRINCE, M. B., 1953, *Phys. Rev.*, **92,** 681.

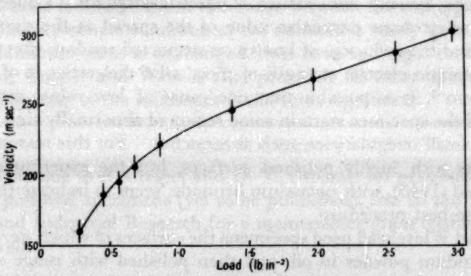
LETTERS TO THE EDITOR

The Variation of Sound Velocity with Stress in Sand

During the course of investigations into the propagation of sound in granular materials, it has been found that the velocity and attenuation of sound in dry sand is markedly dependent upon the load applied to the sand.

In a particular experiment the sand was contained in a 6-inch diameter cylindrical can, lined with $\frac{1}{8}$-inch sponge rubber constituting a pressure-release boundary and isolating the sand acoustically from the wall of the can. A load was applied by weighting a loosely fitting piston containing the barium titanate detector and head amplifier. The can stood on a 3-inch diameter metal disc on the face of the transmitting magnetostriction transducer: in this way the load was being applied in the direction of propagation.

The sound velocity was obtained by pulsing the transmitter output and measuring, by means of an oscilloscope, the pulse transit time for various thicknesses of sand. A family of linear graphs was obtained of sand thickness against transit time for various applied loads: from this was obtained the relationship between velocity and load. One example of the result of such a series of measurements is shown in the figure.



The measurements were taken at 20 kc/s, timing being made to the front of the pulse. Some distortion of the pulse shape was obtained due to the frequency characteristics of the transducers and to the amplifier bandwidths. A further factor introducing distortion is the propagation of complex modes due to beam spread and reflections occurring at the pressure-release wall and to inhomogeneity in the sand. These effects should not greatly affect the velocity measurements since the front of the pulse was timed and the experimental conditions were chosen to minimize the effects and approximate the ideal case of propagation of plane dilatational waves. No appreciable dispersion was noted with these group velocity measurements in the range 10 to 50 kc/s.

The general shape of the curves of sound velocity versus applied load is similar to that obtained theoretically by Gassmann (1951) for the velocity in a close packing of granite spheres at various depths, but there does not appear to be agreement in detail. Gassmann's theory assumes the closest packing of the spheres, porosity 0·26, whereas the sand used in the present measurements has a porosity 0·39. Again Gassmann's result is independent of the sphere radius, whereas investigations carried out here suggest that, even with sands of the same porosity and loading, the velocity is dependent upon the granule size.

It has been observed that the velocity increases, while the attenuation decreases, when loading is applied in the direction of propagation. However, when the load is applied at right angles to the direction of propagation the converse is true. Detailed investigation of this velocity–stress effect and of the effect of stress on attenuation are in progress.

University College of Leicester.
27th February 1956.

E. Matsukawa.
A. N. Hunter.

Gassmann, F., 1951, *Geophysics*, **16**, 673.

The Electric Strength and Surface Condition of the Alkali Halides

When measurements are made of the intrinsic electric strength of solid dielectrics, the results obtained from individual specimens usually exhibit a considerable dispersion, greatly exceeding the known random errors of the measurements. This is because great difficulty is experienced in eliminating the influence of the many spurious effects which can alter the breakdown voltage of a specimen, giving different, usually low, values of electric strength. Consequently it is necessary to select some particular value of the spread as the significant value, bearing in mind the influence of known or suspected random effects.

As the intrinsic electric strength of most solid dielectrics is of the order of 10^6 to 10^7 v cm^{-1}, it is possible that one cause of low values can be that the breakdown of the specimen starts in some region of abnormally high electric stress surrounding small irregularities, such as scratches. For this reason it is usual to use specimens with highly polished surfaces, but the experiments of Austen and Whitehead (1940), with potassium bromide, seem to indicate that this is not necessarily the best procedure.

Austen and Whitehead used specimens the surfaces of which were first ground with carborundum powder in oil, and then polished with rouge on a chamois leather. These specimens gave inconsistent and rather low values of electric strength. When the final polishing was omitted, so that the specimens had a fine matt appearance, the values of electric strength obtained were more consistent and about 50% higher than for the polished specimens.

This result was attributed to the presence in the surfaces of the polished specimens of a few minute scratches, the number and position of which were different between specimens; whereas in the case of the ground surfaces, abnormally high concentrations of electric stress were unlikely because of the screening effect on each other of the close network of scratches.

Some doubt has been thrown on the validity of this explanation by the work of Calderwood, Cooper and Wallace (1953), since no precautions were taken by Austen and Whitehead to ensure that the specimens used contained a minimum amount of strain, and were as nearly alike in this respect as possible. In addition, Austen and Whitehead tested only about five or six specimens of each type, not eliminating therefore the possibility of statistical fluctuations in the results.

Calderwood, Cooper and Wallace found it necessary to anneal the specimens at a temperature near to the melting point, to remove strains introduced in the course of preparation; because they found that the presence of strain in the specimens appreciably increased the spread and average value of the electric strength.

The annealing process renders a previously polished surface translucent. Figure 1 (Plate) is a photomicrograph of a polished surface, and figure 2 (Plate) is a photomicrograph, on the same scale, of a similar surface after annealing, both taken by reflected light. Clearly, a great difference exists in the microscopic condition of the two surfaces, and on the basis of Austen and Whitehead's argument, it would be expected that specimens with surfaces similar to that in figure 2 would have higher and more consistent values of electric strength than polished surfaces.

Twenty-four sodium chloride specimens were prepared, of the usual plane-recessed form, given a high degree of polish on both sides, and then annealed. After annealing, the surfaces were not touched except for measuring, and the application of graphite electrodes. The surfaces were thus similar to that in figure 2. The electric strengths of this group of specimens extended from $0{\cdot}74 \times 10^6$ to $1{\cdot}36 \times 10^6$ v cm^{-1}, and gave an average value of $1{\cdot}04 \times 10^6$ v cm^{-1}.

Previously obtained values for fifty-one specimens polished before and after annealing, that is with surfaces similar to that in figure 1, extended from $0{\cdot}8 \times 10^6$ to $1{\cdot}28 \times 10^6$ v cm^{-1}, and gave an average value of $1{\cdot}03 \times 10^6$ v cm^{-1}.

Clearly, there is no significant difference between the two types of specimen, although the microscopic condition of the surfaces is vastly different. In particular the minimum value is unchanged, thus lending weight to the argument of Calderwood, Cooper and Wallace that provided all other spurious effects are absent a value near to the minimum is the most significant.

ACKNOWLEDGMENTS

Thanks are due to R. Cooper and D. T. Grossart for permission to quote the results for polished specimens (yet to be published), and to the Department of Scientific and Industrial Research for a maintenance grant enabling the work to be carried out.

Electrical Engineering Laboratories,
University of Manchester.
24th May 1956.

W. A. SMITH.

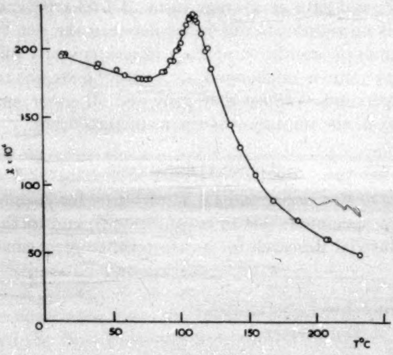
AUSTEN, A. E. W., and WHITEHEAD, S., 1940, *Proc. Roy. Soc.* A, **176**, 33.
CALDERWOOD, J. H., COOPER, R., and WALLACE, A. A., 1953, *Proc. Instn Elect. Engrs*, Pt. IIA, **100**, 105.

Structural Transition and Antiferromagnetism in the Alloy AuMn

Raub *et al.* (1953) have reported that AuMn possesses a CsCl-type lattice above 615°C and an ordered body-centred tetragonal lattice below this temperature. Later, Giansoldati and Linde (1955) observed anomalies in the electrical resistivity–temperature and magnetic susceptibility–temperature curves for an alloy of this composition at 227°c (but none at or near 615°c), the latter taking the form known for antiferromagnetic materials. The structure of the quenched alloy was found to be body-centred tetragonal, in agreement with Raub.

This note describes the preliminary results of an investigation made initially to establish whether the crystallographic change and the possible antiferromagnetism are associated phenomena, as is the case for several other antiferromagnetic materials.

An alloy of composition 49·75 atomic% Mn, prepared *in vacuo* in an alumina crucible from standardized gold (Johnson Matthey, J. M. 70) and vacuum distilled high purity manganese (Hilger H. S. Brand), was examined in powder form in a high temperature x-ray camera, fitted with hemispherical heaters and ring thermocouples of the type described by Basinski and Christian (1954). At high temperatures the photographs showed diffraction lines of a CsCl-type structure and these lines persisted over the temperature range 660°c to 130°c. At 124 ± 2°c, slight blurring of the diffraction lines occurred, whilst at 119 ± 2°c the photograph showed the structure to be ordered body-centred tetragonal. The cubic–tetragonal transformation in this alloy thus occurs near 125°c. At 131 ± 2°c, the cubic unit cell has $a = 3{\cdot}215_5$kx units and at 113 ± 2°c the tetragonal unit cell has $a = 3{\cdot}264_0$kx units and $c/a = 0{\cdot}968$. The cubic structure could not be retained by the most rapid quench.



Magnetic susceptibility per gramme of AuMn as a function of temperature $H = 5000$ Oe.

Susceptibility measurements were made using a Sucksmith ring balance. The specimen was prepared by crushing a small portion of the ingot in an agate mortar in order to avoid pick-up of ferromagnetic impurities, followed by annealing *in vacuo* for 15 hours at 450°c. A powdered specimen was used for the main work to enable direct comparison to be made with the x-ray results. Measurements were made from room temperature up to 232°c, the temperature being measured by means of a fine thermocouple whose junction lay approximately 1 mm below the high-purify aluminium specimen holder. In view of the lack of a calibrating substance of comparable susceptibility, the absolute values of χ in a field of 5000 Oe are estimated to be correct only to 1–2%, but the variation of χ with T is known more accurately. The curve obtained for 5000 Oe is shown in the figure. Similar behaviour was observed at 4000 and 4500 Oe.

Above 70°C, the (χ, T) curve is typical of an antiferromagnetic substance, with a Néel point T_c of 383°K, but above 127°C χ follows the Curie–Weiss law $\chi = C/(T-\theta)$, with $\theta = 359$°K (which is opposite in sign to that of most normal antiferromagnetics) and $C = 7{\cdot}5_1 \times 10^{-3}$. The latter gives a Bohr magneton value of 3·89 per unit cell, suggesting a $3d^7$ state of the Mn atoms if the magnetic moment is associated entirely with them. A martensitic transformation probably occurs as the magnetic interaction must first change sign to account for the peak in the (χ, T) curve, as well as the observed crystallographic distortion. The temperatures at which these two effects occurred appeared not to coincide exactly, although it should be noted that in neither measurement was the thermocouple in actual contact with the specimen.

Below 70°C χ begins to increase again with decreasing temperature. This effect was observed with both a powdered and a polycrystalline lump specimen. Further work is now in progress to study this and the low-temperature region in more detail, but an X-ray photograph taken at liquid-air temperature indicates that no further crystallographic changes occur in the alloy down to this temperature.

We are unable to offer an explanation of the differences between these results and those of the previous workers, unless the behaviour is extremely sensitive to small amounts of impurity and departure from the ideal AuMn composition. In the present work both the X-ray and the magnetic results were obtained with specimens prepared in an identical manner from the same ingot.

We are indebted to Professor E. R. Andrew for providing facilities for this work and to the Council of the Royal Society for a grant to purchase the raw materials. One of us (R.R.P.) would also like to acknowledge a Durham County Post-graduate Research Scholarship.

Physics Department,
University College of North Wales,
Bangor.
16th April 1956.

D. P. Morris.
R. P. Preston.

Basinski, Z. S., and Christian, J. W., 1954, *Proc. Roy. Soc.* A, **223,** 554.
Giansoldati, A., and Linde, J. O., 1955, *J. Phys. Radium*, **16,** 341.
Raub, E., Zwicker, U., and Baur, H., 1953, *Z. Metallk.*, **44,** 312.

Anomalous Lorenz Numbers in Mixed Semiconductors

Goldsmid (1956) has reported anomalously large measured values, for bismuth telluride, of the Lorenz number Λ, which is defined by

$$K = K_1 + \Lambda T \sigma \qquad \ldots\ldots(1)$$

where K and σ are thermal and electrical conductivities respectively and K_1 is the the contribution to K from conduction by lattice modes. He suggests that the anomaly is to be ascribed to transport of energy by ambipolar thermo-diffusion (Price 1954, 1955). The general formula for the Lorenz number of a mixed (two-band) cubic semiconductor with Boltzmann statistics, taking the ambipolar effect into account, is †

$$\Lambda=\left(\frac{\boldsymbol{k}}{\boldsymbol{e}}\right)^2\sigma_1\rho^2\sigma_2\left[\frac{\Delta}{\boldsymbol{k}\mathrm{T}}+3+\gamma_1+\gamma_2\right]^2+\rho\sigma_1\Lambda_1+\rho\sigma_2\Lambda_2, \qquad \ldots\ldots(2)$$

where Λ_1 and Λ_2 are the Lorenz numbers of the separate (conduction and valence) bands, ρ is the resistivity, and the other symbolsare as defined previously(Price 1955). The formulae quoted by Goldsmid (e.g. Eqn (3) of his paper) evidently assume spherical energy surfaces (scalar effective mass) and scattering of carriers by Debye acoustic lattice modes only (for convenience we may speak of these assumptions as defining 'the classical model'). In this case $\Lambda_1=\Lambda_2=2(\boldsymbol{k}/\boldsymbol{e})^2$ and $\gamma_1=\gamma_2=1/2$. Then the bracket $(\epsilon_g/\boldsymbol{k}T+3)$ in his formulae should be $(\epsilon_g/\boldsymbol{k}T+4)$, and the factor 2/3 appearing in his Eqn (3) should be 1/2. With these corrections, one may revise the calculation Goldsmid makes for his Specimen 2 at 300°K, assuming as he does that $\Delta=5{\cdot}8\boldsymbol{k}T$ and that electron and hole concentrations were equal, but taking the mobility ratio as 2, the value he quotes. Then for the classical model the ratio of the predicted value of Λ to its 'normal' value of $2(\boldsymbol{k}/\boldsymbol{e})^2$ is

$$1+\frac{1}{2}\left[\frac{2\times 1}{(2+1)^2}\right](5{\cdot}8+4)^2=11{\cdot}7.$$

This result compares with Goldsmid's experimental value of 13 and his calculated value of 14. Allowing for the uncertainties in the data, and the assumptions made (Boltzmann statistics and the classical model),‡ the agreement of the present estimate with experiment is at least as good as might be expected, if not coincidentally close. It should be added that the theory of electronic thermal conductivity leading to (2) omits a further possibly important contribution : that from the energy transported by the stream of excitons which would be expected to flow from the hot end to the cold end of the crystal. This latter contribution should be independent of (moderate) doping, like the lattice term K_l, and would be expected to vary with temperature roughly as $\exp(-\Delta/\boldsymbol{k}T)$. Unlike the contribution represented by (2), it should not contribute to the Righi–Leduc effect, or to the Nernst effect (Price 1955), and should not exhibit 'magnetoresistance' (decrease in a strong magnetic field). It should be possible to determine the contribution to K represented by (2) by plotting a 'doping curve' of K *versus* resistivity or Hall mobility, at constant temperature, for a series of crystal specimens. For the doping range where Δ and the γ's and Λ's (and in the second case the band Hall mobilities) do not differ from their values for a pure perfect crystal, equation (2) predicts definite forms for the doping curves. By fitting the points to these, one may expect to determine Λ_1, Λ_2 and the square bracket of equation (2). (The latter magnitude may be determined independently from a doping curve for the Nernst coefficient (Price 1956), and also from one for the thermoelectric power.) This procedure if successful would allow the part of the experimental thermal conductivity given by (2) to be subtracted, so that the temperature dependence of the remainder could be examined for indications of an exciton contribution. For a non-cubic crystal, such as bismuth telluride, the scalar formula (2) applies only for conduction in special directions. In a general direction Λ is a tensor, and is defined by the obvious generalization of (1) as proportional to the contracted product of the tensors ρ and $K-K_l$. The formula (2) is correctly generalized when $\Delta/\boldsymbol{k}T+3$ is multiplied by the unit tensor; when ρ, the σ's, γ's and Λ's are tensors; and when products of tensors are contracted.

---

† This formula is more complete than given previously, in that the parameters α and β are here given in terms of γ_1 and γ_2, because I have since shown that (as a consequence of Onsager's reciprocal relations) the thermal energy transported with a drifting electron or hole is $(\gamma_1+3/2)\boldsymbol{k}T$ or $(\gamma_2+3/2)\boldsymbol{k}T$ respectively.

‡ The 'experimental' value of Λ deduced by Goldsmid, by comparing K at 150° with K at 300°, itself depends on the assumption that $\Lambda_1=2(\boldsymbol{k}/\boldsymbol{e})^2$ at 150°, and might be wrong by of order 10% on that account. The assumption of spherical energy surfaces may be considered dubious for the Bi_2Te_3 crystal, which is not cubic. It is not stated by Goldsmid whether the experimental value of Δ (i.e. ϵ_g) which he quotes is thought to correspond to the absorption edge for direct transitions, in which case it would exceed the true forbidden energy gap except in special conditions, or to that for indirect transitions.

As Goldsmid notes, Fröhlich and Kittel (1954) independently suggested the ambipolar contribution to the electronic thermal conductivity as the explanation of the anomalously large values of the Lorenz number for indium antimonide found by Busch and Schneider (1954). The experimental values are, however, much too large to be explained in this way (Busch and Schneider 1954, Price 1955). An estimate of the exciton–drift contribution, using values of the coefficient of diffusion given by the theory of Ansel'm and Firsov (1955) for the scattering of excitons by phonons, indicates that—at least if one counts only the ground-state excitons—this contribution to K is probably of the same order of magnitude as the ambipolar one, and thus too small (the data are too uncertain for one to be more precise). On further consideration it is evident that this exciton explanation in any case would have had no foundation. The Bohr radius of the hydrogenic ground state of the exciton in indium antimonide is about 700 Å. The product of the volume of an exciton and the estimated exciton concentration (at 300°K, say) is then a large number. Hence the basis of the estimation of the concentration, the neglect of the energy of interaction between excitons, is unwarranted. The explanation might have been an attractive one because it has been found (Busch 1955, private communication) that the 'magneto-resistance' effect for the thermal conductivity is zero, or at least small compared with that for the electrical conductivity. A different possibility is suggested by this difficulty in the exciton explanation: if incipient excitons interact drastically, then the true state of the electron–hole assembly may presumably be likened to that of a quantum liquid. Although doubtless the self-consistent-field principle applies to a first approximation, the exciton level being quenched, the assembly should support a high density of plasma modes (Pines 1955) in which the oscillations of electron density and hole density are in phase. The anomalous heat flow may then be transported by travelling plasma mode quanta. This suggestion is of course no more than speculation, until and unless it is supported by adequate experimental evidence or by sufficient knowledge of the mean free paths of plasmons in these conditions. It may, however, suggest fresh approaches to the experimental study of indium antimonide.

Watson Laboratory,
Columbia University,
New York City.
9th April 1956.

P. J. PRICE.

Ansel'm, A. I., and Firsov, Iu. A., 1955, *J. Exper. Theor. Phys. USSR*, **28**, 151; *Soviet Physics JETP*, **1**, 139.
Busch, G., and Schneider, M., 1954, *Physica*, **20**, 1084.
Fröhlich, H., and Kittel, C., 1954, *Physica*, **20**, 1086.
Goldsmid, H. J., 1956, *Proc. Phys. Soc.* B, **69**, 203.
Pines, D., 1955, *Solid State Physics*, **1**, 367.
Price, P. J., 1954, *Phys. Rev.*, **95**, 596; 1955, *Phil. Mag.*, **46**, 1192; 1956, *Phys. Rev.*, in the press.

The Airy Pattern in Systems of High Angular Aperture

Workers in light optics and microwave optics have been suspecting for some time that when the angular aperture of the image forming pencil in an aberration free optical system is sufficiently increased, the diffraction image undergoes a more substantial modification than that of a simple diminution predicted by the classical formula of Airy. This possibility is of considerable practical interest in connection with resolving power in wide aperture systems and has involved much speculation†.

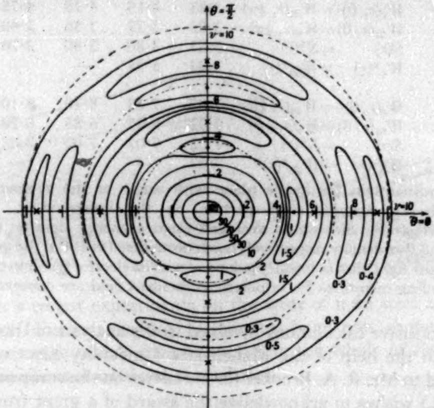
The difficulties which present themselves in the theoretical analysis of this problem are that the usual scalar treatment is not applicable when the angular aperture exceeds a moderate value, and that a rigorous boundary value treatment based on Maxwell's theory requires the specification of exact boundary conditions in the plane of the aperture—conditions which are practically unobtainable. Airy's solution was extended by Hopkins (1943) to systems with semi-apertures up to 40°–50°, where it became apparent that the effects of polarization give rise to some modification in the pattern, the investigation of which would require more refined techniques.

We have recently succeeded in finding a solution to this problem, valid for systems of any angular aperture, in which (as is always the case in light optics) the distance between the diffracting aperture and the geometrical focus is large compared with the wavelength; the image field may then be considered not as one that is specified by its boundary values, but rather as one which must have prescribed asymptotic behaviour with increasing distance from the geometrical focus. Such a solution can be obtained with the help of Luneberg's diffraction integrals (1944), which are essentially a vectorial generalization of a well-known representation, due to Debye (1909), of a diffracted pencil of converging rays. The full mathematical analysis and the results obtained will be published at a later date; here only the main conclusions will be presented.

We have considered an aplanatic optical system with object at infinity and studied the case of linearly polarized, as well as the case of unpolarized, light incident upon the entrance pupil. When the light is linearly polarized, the lines of constant time averaged electric energy density in the image plane are found to depart, with increasing semi-aperture α, more and more from circular form, being approximately elliptical in the central part of the image and having more complex structure elsewhere (see figure, which illustrates the case $\alpha = 60°$). The

† See for example the discussion by Toraldo di Francia, Osterberg, Zernike and Maréchal, reported in the *Proceedings of Symposium on Optical Image Evaluation* (1954).

major axes of the ellipses are in the direction of the electric vector of the incident wave; thus, *with linearly polarized light and using detectors of electric energy (e.g. a photographic plate) our solution predicts an increase in resolving power in wide aperture aplanatic optical systems for measurements in directions at right angles to the electric vector of the incident wave.* This result is in agreement with the conclusions of Hopkins (1943, 1945), and appears to be supported by early experiments recorded by Carpenter (1901). Further, our analysis shows that the lines of constant magnetic energy density are identical with those of the electric energy density but are rotated through 90° about the optical axis. The lines of constant energy flux are found to be circles whose radii are practically identical (for any α) with those of the Airy intensity pattern.



Contour lines of time averaged electric energy density W_e, in the focal plane of an aplanatic system of angular semi-aperture $\alpha=60°$, the incident light being polarized with its electric vector in the direction $\theta=0$. The values are normalized to 100 at the geometrical focus. The dotted lines represent the dark rings as given by Airy's formula. Axial minima are denoted by dots and axial maxima by crosses.

When the incident light is *unpolarized*, the electric and magnetic energy densities are found to be equal. Naturally the contours are now circles, *but their radii are found to be somewhat larger (of the order of* 8% *when* $\alpha=60°$) *than the corresponding radii as calculated from the Airy formula*; also the energy density functions no longer have strict zero values, and have no pronounced subsidiary maxima or minima.

When the angular aperture is very small ($\alpha\to 0$) the images in polarized and unpolarized light are identical and the time averaged energy densities and flux are equal to each other (see table), and are found to be given by the intensity formula of Airy.

Finally, it must be mentioned that the basic integrals of our theory are identical with those given by Ignatowsky (1919) in the analysis of the same problem many years ago. Our conclusions, however, go considerably further than those derived in his valuable, but little known, paper.

Data showing the Structure of the Electromagnetic Diffraction Image in the Focal Plane of an Aplanatic System

Value	Quantity	$v=(2\pi/\lambda)r\sin\alpha$				
		$\alpha\to 0$	$\alpha=30°$	$\alpha=60°$	$\alpha=75°$	$\alpha\to 90°$
80	$W_e(r,0)=W_m(r,\frac{1}{2}\pi)$	0·93	0·98	1·16	1·33	1·48
	$W_m(r,0)=W_e(r,\frac{1}{2}\pi)$	0·93	0·93	0·89	0·86	0·84
	$S(r)=S^*(r)$	0·93	0·93	0·93	0·93	0·93
	$W_e^*(r)=W_m^*(r)$	0·93	0·95	1·00	1·03	1·05
50	$W_e(r,0)=W_m(r,\frac{1}{2}\pi)$	1·60	1·70	2·00	2·22	2·37
	$W_m(r,0)=W_e(r,\frac{1}{2}\pi)$	1·60	1·60	1·54	1·50	1·45
	$S(r)=S^*(r)$	1·60	1·60	1·60	1·60	1·57
	$W_e^*(r)=W_m^*(r)$	1·60	1·64	1·75	1·80	1·84
1st Min.	$W_e(r,0)=W_m(r,\frac{1}{2}\pi)$	3·83	4·15	4·75	4·75	4·70
	$W_m(r,0)=W_e(r,\frac{1}{2}\pi)$	3·83	3·77	3·56	3·40	3·29
	$S(r)=S^*(r)$	3·83	3·80	3·80	3·76	3·70
	$W_e^*(r)=W_m^*(r)$	3·83	3·95	—	—	—
2nd Min.	$W_e(r,0)=W_m(r,\frac{1}{2}\pi)$	7·02	7·03	8·10	8·10	8·00
	$W_m(r,0)=W_e(r,\frac{1}{2}\pi)$	7·02	6·95	6·85	6·70	6·55
	$S(r)=S^*(r)$	7·02	7·02	7·02	7·02	6·95
	$W_e^*(r)=W_m^*(r)$	7·02	—	—	—	—

(r,θ)=polar coordinates in the image plane with origin at the geometrical focus; λ=wavelength; α=angular semi-aperture on the image side; $v=(2\pi/\lambda)r\sin\alpha$; $W_e(r,\theta)$, $W_m(r,\theta)$, $S(r)$=time averaged electric energy density, magnetic energy density, and flux density respectively (each normalized to 100 at the geometrical focus), with incident light that is linearly polarized with its electric vector in the direction $\theta=0$. Corresponding quantities for unpolarized incident field are denoted by an asterisk.

The very extensive calculations involved in the analysis of this problem were carried out with the help of the Manchester University Electronic Computer. We are indebted to Mr. R. A. Brooker for his advice on the computational aspects. One of us (B.R.) wishes to acknowledge the award of a grant from the National Research Development Corporation; the other (E.W.) is indebted to Manchester University for the award of an Imperial Chemical Industries Research Fellowship, during the tenure of which this work was carried out.

Computing Machine Laboratory, University of Manchester. B. RICHARDS.
The Physical Laboratories, University of Manchester. E. WOLF.

12th June 1956.

CARPENTER, W. B., 1901, *The Microscope and its Revelations* (London: Churchill), 8th Edn., p. 381.
DEBYE, P., 1909, *Ann. Phys., Lpz.*, (4), **30**, 755.
HOPKINS, H. H., 1943, *Proc. Phys. Soc.*, **55**, 116; 1945, *Nature, Lond.*, **155**, 275.
IGNATOWSKY, B. C., 1919, *Trans. Opt. Inst., Petrograd*, Vol. I, Paper IV.
LUNEBERG, R. K., 1944, *Mathematical Theory of Optics* (Mimeographed lecture notes; Brown University, Providence, R.I.)
SYMPOSIUM ON OPTICAL IMAGE EVALUATION, 1954, *Nat. Bur. Std., Wash.*, Circular 526, pp. 35–37.

The Relation between Wear and Distance Run, for Steel and Coated Steel Surfaces

It is well known that when two surfaces of a bearing are run together, at not unduly heavy normal pressure, they become progressively smoother and better matched in shape, i.e. they become run-in. Measurements of the wear during this process are known to be of the type shown in figure 1, having an initial stage of rapid wear, giving place to a later lower rate. It appears, however, that no adequate quantitative relationship between the amount of wear W and the distance run D has previously been recognized and that experiments repeated under similar conditions show a large variation in the amount of wear observed.

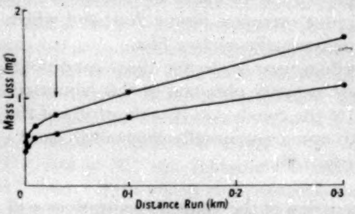


Figure 1

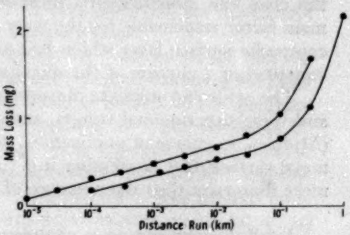


Figure 2.

The writer's recent experiments on the wear of mild steel against mild steel under an inert lubricant (medicinal paraffin) have now shown (i) that over a relatively large range which covers practically the whole of the running-in, i.e. for $D = 1$ cm to about 5×10^3 cm, the relation

$$W = c \log kD \qquad \ldots\ldots(1)$$

where c and k are constants, is followed with high accuracy, as is seen in figure 2, and (ii) that although in repeated experiments k varies, nevertheless c is remarkably constant, as figure 2 also shows. For comparison, and to illustrate the variation from experiment to experiment in the actual wear, figure 1 represents the same data as in figure 2, plotted on the more usual linear scale, and the arrow in figure 1 indicates the distance D up to which equation (1) is followed (cf. figure 2).

In these experiments three discs, 0·635 cm in diameter at the face, were run against a ring of internal and external diameters 6·19 and 8·10 cm respectively. Surfacing by abrasives was avoided, the discs being finely lathe-turned using a standard cut of 0·01 mm pitch, and the ring was machined so that the surfacing marks (coarser than on the discs) were substantially radial. The discs were mounted on a steel plate in such a way that their turning marks were perpendicular to the median circumference of the ring, i.e. perpendicular to the direction of relative motion. The relative speed of motion of ring and discs was 28·4 cm sec^{-1}, and the nominal load 100 kg cm^{-2}. The wear mass-loss of the three discs was determined by weighing on a semi-microbalance.

The wear of mild steel discs coated with various iron compounds was studied under the same conditions as above. Layers of FeS and Fe_3O_4 a few microns thick were found to obey the logarithmic law of equation (1), each film having a characteristic *c* value, which for Fe_3O_4 was greater than for mild steel, and for FeS was still larger.

Layers of α-Fe_2O_3 (a few microns thick) on mild steel did not obey the logarithmic relation, clearly because of appreciable pick-up of metal from the opposing steel ring, as was indicated by the initial small *increase* in weight of the discs.

Mild steel rubbing on mild steel, but lubricated with liquid paraffin containing 0·5% diethyl trisulphide instead of the pure liquid paraffin used normally, also obeyed the logarithmic law, and had a value of *c* intermediate between that for FeS and that for untreated mild steel.

The wearing surfaces were examined in the microscope, and the wear debris was examined by electron diffraction and electron microscopy. Together with the mass-loss measurements these observations led to the conclusion that the main factor responsible for the wear properties was probably the formation of a composite surface layer which had undergone extensive plastic flow, and which consisted of a mixture of the metal and non-metallic surface film.

The oxide and sulphide films do not reduce wear under the above conditions, and these experimental results, and others recently obtained in this laboratory (Midgley, in course of publication) point to the conclusion that the shape of the metal surface, that is, whether it is smooth or microscopically channelled, etc., is more important than the presence of a surface film.

A full account and more complete discussion of the above investigations will be published in due course. The research formed part of the programme of the Department of Scientific and Industrial Research Mechanical Engineering Research Board, and this communication is made by permission of the Director of Mechanical Engineering Research, Department of Scientific and Industrial Research. The author would also like to thank Dr. H. Wilman for many helpful discussions.

Applied Physical Chemistry Laboratory,
Department of Chemical Engineering,
Imperial College, London.
23rd April 1956.

M. R. PIGGOTT.†

---

† Now at Research Laboratories of The General Electric Co. Ltd., Wembley, Middx.

Paramagnetic Resonance of Uranium Ions

Paramagnetic resonance has been observed at 20°K in single crystals of CaF_2 and SrF_2 containing small quantities of uranium and neodymium as impurities. The CaF_2 crystals were grown by the Stockbarger method in a vacuum furnace from powdered, colourless, natural fluorite previously mixed with about 0·1% by weight of uranium or neodymium fluoride; a similar method was used for the SrF_2 crystals but the percentage was about 1%. In each case the resonance spectrum shows three types of magnetic ion each with axial symmetry about one of the cube axes. The values of $g_{\parallel}$ and $g_{\perp}$ are:

	Uranium		Neodymium	
	$g_{\parallel}$	$g_{\perp}$	$g_{\parallel}$	$g_{\perp}$
CaF_2	3·501	1·866	4·412	1·301
SrF_2	3·433	1·971	4·289	1·505

(errors : $\pm 0{\cdot}008$ in $g_{\parallel}$; $\pm 0{\cdot}002$ in $g_{\perp}$).

These results show that in each case the magnetic ion has an odd number of electrons, and hence must be in an odd valency state. The only plausible assumption is therefore of U^{3+} and Nd^{3+} ions; in the latter case this is confirmed by measurements of the hyperfine constants of ^{143}Nd, ^{145}Nd, which bear the same ratio to the *g*-values as observed previously for Nd^{3+} ions. Some lines, probably due to hyperfine structure of the rare isotope ^{235}U, were observed, but these could not be identified for certain because of other impurity lines.

CaF_2, SrF_2 crystals belong to the space group O_h^5; the structure may be regarded as a simple cubic lattice of fluorine ions, with divalent cations at the alternate body centres. We have considered the following models of the structure when a trivalent cation is introduced: (*a*) M^{3+} replaces M^{2+}, and an excess F^- ion occupies the interstitial position at the centre of the adjacent cube of fluorine ions: (*b*) M^{3+} replaces M^{2+}, and an O^{2-} ion replaces an F^- ion; (*c*) M^{3+} replaces M^{2+}, and an M^+ ion replaces a second M^{2+}; (*d*) two M^{3+} ions replace three M^{2+} ions. If the charge compensating ions are on adjacent sites, models (*b*) and (*c*) would lead to spectra with four types of ions each with symmetry about a trigonal axis. Model (*d*) would probably not give axial symmetry, and strong interaction between the two M^{3+} ions would be expected if they are close together. Model (*a*) would lead to the observed type of spectrum, and it is the only one favoured by the x-ray and density measurements of Zintl and Udgård (1939) on similar crystals with large quantities of trivalent ions.

The interstitial fluorine ion in model (*a*) reduces the symmetry at the cation site from cubic to tetragonal, the tetragonal axis being one of the three cubic axes. If the configuration of a U^{3+} ion is $5f^3$, then its ground state (by Hund's rules) should be $^4I_{9/2}$, the same as $Nd^{3+}(4f^3)$. In a field of four-fold symmetry, the ground state $J=9/2$ would be split into a number of doublets, whose states are characterized by admixtures of different values of J_z either of the form

$$\cos\theta\,|\pm 5/2\rangle + \sin\theta\,|\mp 3/2\rangle$$

or

$$\alpha\,|\pm 9/2\rangle + \beta\,|\pm 1/2\rangle + \gamma\,|\mp 7/2\rangle],\quad (\alpha^2+\beta^2+\gamma^2=1).$$

The high values of $g_{\parallel}$ for Nd cannot be fitted by the first of these admixtures, but all the results can be fitted by the second because of the additional parameter. The values of γ required are very small, and if we put $\gamma=0$, we have the relation $g_{\parallel}+8g_{\perp}/5=9\Lambda=6{\cdot}54$, where $\Lambda=8/11$ is the ordinary Landé *g*-factor for a free $^4I_{9/2}$ ion. The experimental values of $g_{\parallel}+8g_{\perp}/5$ all lie between 6·49 and 6·70, so that it appears likely that the ground state is a Kramers doublet of the second type in all cases. Work on other lanthanide ions is in progress, from which more detailed information on the crystal field should be obtained.

An interesting feature of the U^{3+} spectrum is the presence of a complex hyperfine structure due to the fluorine nuclei. In certain directions where this hyperfine structure is well resolved an even number of components is observed, and this can only arise from interaction with an odd number of fluorine nuclei. There must be a contribution from the interstitial F^- ion comparable with that from the eight F^- ions on normal lattice sites round the uranium ion. No fluorine hyperfine structure is observed on the neodymium resonance, confirming that the 4f electrons take no part in the chemical binding of the ion, while the 5f electrons, with their more extended wave functions, have a small overlap on to the fluorines.

We are indebted to Professor R. V. Jones for making the crystal-growing facilities available to us.

†Clarendon Laboratory,
University of Oxford.
‡Department of Natural Philosophy,
Aberdeen University.
1st May 1956.

B. Bleaney.†
P. M. Llewellyn.†
D. A. Jones.‡

Zintl, E., and Udgård, A., 1939, *Z. anorg. Chem.*, **240**, 150.

Some Electrical Characteristics of Single Crystal Iron Monotelluride

In 1924 Brukl prepared what was probably tetragonal iron monotelluride by reducing ferrous salts with hydrogen ditelluride, but no electrical measurements of this phase have so far been reported. Large single crystals have been prepared for magnetic measurements, and the temperature variation of the electrical properties will be described.

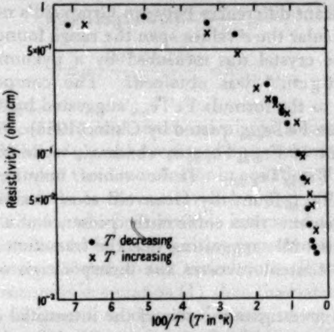
The iron and tellurium (containing less than 0·0005% impurities as shown by spectrographic analysis) were mixed in the required proportions and sealed off in a silica tube at a pressure of 10^{-6} cm Hg. The temperature was raised to 920°c and held for 36 hours. Cooling in a temperature gradient then took place over a period of 17 hours. The composition of the ingot produced in this way corresponded to $FeTe_{0\cdot933}$, and x-ray and optical examination showed that the lower centimetre of the ingot consisted of a single crystal of the β phase, and above it a mixture of β with the orthorhombic ditelluride (ϵ). Grown in this way the β section should be typical of the tellurium rich end of the phase.

A specimen of dimensions 6·38 mm × 1·9 mm × 0·38 mm was cleaved in the (001) plane from the single crystal 0·4 cm from the bottom of the ingot. Contact to the crystal was made by means of a weld of fine copper wire. Current leads at each end and two leads at each side were arranged to allow two resistivity and two Hall effect measurements by a standard potentiometer method. The resistance measurements were made in the [100] direction.

It proved impossible with the apparatus available to obtain a reliable Hall effect measurement but the number of conduction electrons was of the order of 10^{20} cm^{-3} or greater. No change could be detected at low temperature; however the resistance showed an unusual variation with temperature, increasing by a factor of 50 between room temperature and 40°K, and thereafter changing little down to liquid helium temperatures. It will be seen from the figure that the curves obtained on cooling and heating are not identical and in particular the room temperature resistance has increased by some 20% as a result of the cycle.

Similar changes were observed on subsequent runs, and these minor variations may be attributed to thermal stresses in the specimen.

The curve of points in the figure suggests semiconductor behaviour. An increase in resistivity may be due either to a decrease in the number of conduction electrons or to a decrease in the mobility. Preliminary measurements on other samples show that the substance is not an intrinsic semiconductor. If it were an extrinsic semiconductor with a decreasing electron concentration, then the Hall effect should have been easily measurable at low temperatures. Hence it seems likely that the resistivity behaviour is governed by mobility changes. A somewhat similar resistance change in selenium has been reported by Plessner (1951), although in selenium the number of conduction electrons is very much lower (10^{14}). Plessner found that a fairly good straight line plot of ρ against $1/T$ was obtained for the restricted temperature range investigated, the number of conduction electrons remaining constant: the presence of intergranular barriers was suggested as an explanation. For iron telluride the plot of log ρ against $1/T$ (see figure) was not linear and so an explanation by means of a mixture of potential barriers of differing heights must be somewhat arbitrary. When ρ was plotted against T on a logarithmic scale for this iron telluride, a fairly good straight line was obtained between room temperature and 40°K with a slope of about 1·8, but this cannot be taken to imply scattering by ionized impurities since the substance, if it is a semiconductor, is certainly very degenerate.



In view of the uncertainty as to the conduction mechanism it was thought desirable to examine the crystal for the presence of interstitial atoms; lattice parameters and densities were therefore found for different levels in the ingot with Fe radiation and a 19 cm diameter powder camera. Successive extrapolations of the parameters against $\frac{1}{2}(\cos^2\theta/\sin\theta+\cos^2\theta/\theta)$ with adjustments of the ratio a/c were made, and the results are shown in the table.

Material	a(Å)	c(Å)
Bottom of single crystal	3·8215 ± 0·0005	6·2695 ± 0·0005
Top of single crystal	3·8215 ± 0·0005	6·2704 ± 0·0005
From mixed layer	3·829 ± 0·002	6·285 ± 0·002
Fe rich†	3·823$_0$ ± 0·001	6·276$_7$ ± 0·001
Te rich†	3·819$_8$ ± 0·001	6·280$_5$ ± 0·001

† Gronvold *et al.* (1954).

It will be seen that no change can be detected between the top and bottom of the single crystal, but there is a difference for material from the mixed β–ϵ layer. The only data with which a comparison can be made are those of Gronvold *et al.* (1954), on specimens quenched from 300°c. The values quoted are shown in the table for both the iron and tellurium rich side of the β phase. It will be seen that there are significant differences between Gronvold's material and that being described. In particular the c values span the range found by Gronvold. The density of the single crystal was measured by a pyknometer method, and a value of $6{\cdot}98 \pm 0{\cdot}06\,\mathrm{g\,cm^{-3}}$ was obtained. The composition of the single crystal will be near to the formula $\mathrm{FeTe_{0\cdot9}}$ suggested by Gronvold *et al.* (1954) and certainly less than $\mathrm{FeTe_{0\cdot95}}$ quoted by Chiba (1955). If a value of $\mathrm{FeTe_{0\cdot90}}$ is used a cell content of $\mathrm{Fe_{2\cdot23}Te_{2\cdot04}}$ is obtained, while if $\mathrm{FeTe_{0\cdot95}}$ is assumed the cell content is $\mathrm{Fe_{2\cdot18}Te_{2\cdot06}}$. These values may be compared with the cell content of $\mathrm{Fe_{2\cdot19}Te_{1\cdot98}}$ found by Gronvold *et al.* for the iron rich material. The present measurements thus confirm the presence of a high concentration of interstitial iron atoms while suggesting that the transition across the phase field to higher tellurium content involves the incorporation of a small number of interstitial tellurium atoms.

It is proposed to investigate the role of the interstitial atoms on the electron scattering by preparing a series of specimens with different compositions.

Department of Natural Philosophy,
Marischal College,
Aberdeen.
13th June 1956.

D. M. Finlayson.
D. Greig.
J. P. Llewellyn.
T. Smith.

Brukl, A., 1924, *Monatshefte*, **45**, 471.
Chiba, S., 1955, *J. Phys. Soc. Japan*, **10**, 837.
Gronvold, F., Haraldsen, H., Vihovde, J., 1954, *Acta Chem. Scand.*, **8**, 1927.
Plessner, K. W., 1951, *Proc. Phys. Soc.* B, **64**, 671.

The Hall Effect in Semiconductors

Approximate formulae have been derived for the Hall constant of non-degenerate semiconductors for small magnetic fields when the scattering of the charge carriers, assumed to be of one sign, arises from ionized impurity centres (Shockley 1950) and from a combination of impurity centres and lattice vibrations (Johnson and Lark–Horovitz 1951, Jones 1951). The object of this letter is to examine the approximations used.

The expression for the Hall constant can be written as follows:

$$R = \frac{3\sqrt{\pi}}{4nec} \frac{\int_{\eta}^{\infty} \tau^2 \eta^{3/2} \exp(-\eta)\, d\eta}{\left[\int_0^{\infty} \tau \eta^{3/2} \exp(-\eta)\, d\eta\right]^2} \qquad \ldots\ldots (1)$$

where h is the concentration of charge carriers, τ the relaxation time and η is the reduced chemical potential.

Impurity scattering.

For the case of impurity scattering the relaxation time can be written as follows (Mansfield 1956):

$$\tau = \frac{\sqrt{(2m^*)}\epsilon^2(\eta \boldsymbol{k}T)^{3/2}}{\pi n \boldsymbol{e}^4 f(x)},$$

$$f(x) = \log(1+x) - \frac{x}{1+x},$$

$$x = \frac{2m^*\epsilon(\boldsymbol{k}T)^2\eta}{\hbar^2\pi \boldsymbol{e}^2 n} = a\eta,$$

and ϵ = dielectric constant.

Values of the Hall constant have been calculated assuming that $f(x)$, which is a slowly varying function of η, can be replaced by a constant using the value of η which makes the integrand of the denominator in equation (1) a maximum ($\eta = 3$). However, it is not correct to use the same value of $f(x)$ in the numerator of the equation since the value of η which makes this integrand a maximum is 4·5 and the values of $f(x)$ which should be used for this approximate derivation of R are different for the two integrals and do not cancel out. If the appropriate values of $f(x)$ are used in the integrals in equation (1), then the Hall constant is dependent on a where $x = a\eta$. Figure 1 shows the values of $r = Rn\boldsymbol{e}c$ as a function of a obtained by this approximate method. It is obvious that when a is small the approximate evaluation of the integrals is not correct. Thus for $a \to 0$ $f(x) \to \frac{1}{2}(a\eta)^2$ and $r = 1{\cdot}178$. It has been necessary therefore to evaluate the integrals numerically, and figure 1 also shows the correct variation of r with a.

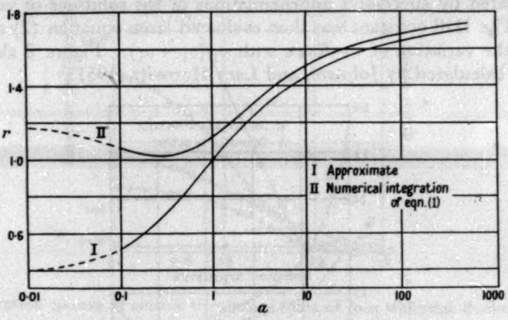


Figure 1. I, approximate ; II, numerical integration of equation 1.

One should remember that the theory of impurity scattering is based on the Born approximation and when a becomes small the validity of this approximation is doubtful.

Combination of impurity and lattice scattering.

When both lattice and impurity scattering occurs then the relaxation time for lattice scattering can be written as $\tau_{\mathrm{L}} = C_{\mathrm{L}}\eta^{-1/2}$ where C_{L} is a constant independent of η and

$$\tau = \frac{C_{\mathrm{L}}\eta^{3/2}}{\eta^2 + \alpha}$$

$$\alpha = \frac{\sigma_{\mathrm{L}}}{\log(1+\exp\eta^*)}\,\frac{3\boldsymbol{h}^3 n \boldsymbol{e}^2}{32\epsilon^2 m^*(\boldsymbol{k}T)^3}f(x).$$

For interpreting experimental results in this case, little is gained by treating $f(x)$ as a constant because of the the difficulty of calculating the appropriate values of α to be used in the integrals which determine the Hall constant.

A comparison, however, was made between the results obtained by Johnson and Lark-Horovitz (1951) and those obtained when appropriate values of α are calculated as follows.

Thus

$$R=\frac{3\sqrt{\pi}}{4nec}\frac{\displaystyle\int_0^\infty \frac{\eta^{9/2}}{(\eta^2+\alpha_1)^2}\exp(-\eta)d\eta}{\left[\displaystyle\int_0^\infty \frac{\eta^3\exp(-\eta)d\eta}{(\eta^2+\alpha_2)}\right]^2} \qquad \ldots\ldots(2)$$

with

$$\alpha_2=6\frac{\sigma_L}{\sigma_I} \qquad \alpha_1=\alpha_2\frac{f(a\eta_1)}{f(a\eta_2)}$$

and where η_1 and η_2 are the values of η which make the integrands of equation (2) a maximum and are determined by the equations

$$\eta_2^3-\eta_2^2+\alpha_2\eta_2-3\alpha_2=0, \qquad \ldots\ldots(3)$$

$$\eta_1^3-\tfrac{1}{2}\eta_1^2+\alpha_1\eta_1-4{\cdot}5\alpha_1=0. \qquad \ldots\ldots(4)$$

It is possible to introduce the parameter σ_L/σ_I where σ_L and σ_I are the lattice and impurity conductivity because the integral in the denominator is the same as that which determines the conductivity. However, it must be remembered that σ_I is not simply the conductivity which would occur if impurity scattering only occurred but must be determined by the method outlined by Mansfield (1956).

Assuming a value of $a=10$, various values of σ_L/σ_I were chosen and α_1 was then calculated by successive approximations of the solutions of equations (3) and (4). The Hall constant was then evaluated from equation (2) and figure 2 illustrates the variation of $r=Rnec$ with $\sigma_I/(\sigma_I+\sigma_L)$. Figure 2 also gives the values of r calculated by Johnson and Lark-Horovitz (1951).

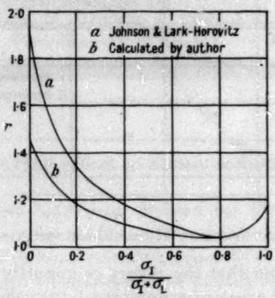


Figure 2.

The analysis given in this letter should also be applied to the other electrical properties of semiconductors. In the case of magneto-resistance, the modification introduced should be relatively important.

Bedford College,
London.
1st June 1956.

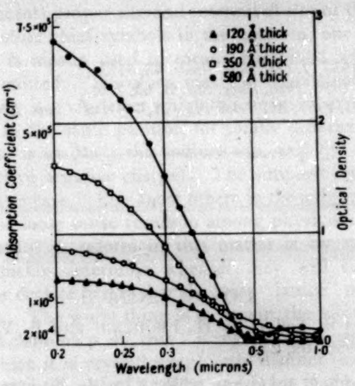
R. Mansfield.

Johnson, V. A., and Lark-Horovitz, K., 1951, *Phys. Rev.*, **82,** 977.
Jones, H., 1951, *Phys. Rev.*, **81,** 149.
Mansfield, R., 1956, *Proc. Phys. Soc.* B, **69,** 76.
Shockley, W., 1950, *Electrons and Holes in Semiconductors* (New York : D. Van Nostrand).

The Absorption Spectrum of Solid Arsenic Trisulphide

Optical absorption in thin films of arsenic trisulphide has been measured as part of a study of the properties of some sulphides of arsenic, antimony and bismuth.

Pure arsenic trisulphide was prepared by treating a solution of arsenious oxide in dilute hydrochloric acid with hydrogen sulphide. Films up to 600Å thick were prepared by vacuum evaporation from a tungsten strip on to fused quartz discs. The absorption of a film was determined in a Beckman Model DU spectrophotometer by measuring the optical density of the specimen and quartz disc mount relative to that of a similar quartz disc alone. The figure shows the optical density as a function of wavelength from 0·2 to 1 μ for four films of different thicknesses. The thickness of the thickest film was determined approximately by weighing and hence absolute values of the absorption coefficient were calculated; these values were used to calculate the thickness of the other films. In the figure, values of absorption coefficient are shown alongside those of optical density. The spectra are reproducible except that in the absorption spectrum of the thickest film there is a marked dip at about 0·5 μ. Similar dips at the edge of the lattice absorption have been reported for other substances (Gibson 1950) and result from slight variations in the method of preparation of the layers.



The absorption spectra of arsenic trisulphide films of four different thicknesses.

The existence of significant true absorption beyond 0·5 μ is questionable. The transmission of plates of vitreous arsenic trisulphide, from 0·02 to 0·30 cm thick, has been measured from 0·5 to 15 μ (Frerichs 1953); the difference in transmission of two plates of different thickness gives an absorption coefficient of about 1 cm^{-1} at 1 μ. Thus in the evaporated films the apparently large value, about 10^4 cm^{-1}, of the absorption coefficient beyond 0·5 μ arises from reflection and scattering losses. The small absorption of the thinnest film at 1 μ shows that reflection losses are negligible in this region and that the apparent absorption in the thicker films is due almost entirely to scattering.

The main features of the spectrum are the sharp rise at 0·5 μ, the continuous increase to 0·2 μ and the magnitude of the absorption coefficient at this wavelength. The absorption coefficient at 0·2 μ is so large that the absorption must be attributed to the bulk lattice and not to impurities. The threshold of the lattice absorption is 0·5 μ and the photoconductivity threshold lies between 0·430 and 0·557 μ (Rekalova 1951). Thus photoconductivity is associated with lattice absorption.

The author is indebted to Dr. J. A. Kitchener of Imperial College, London, for many helpful suggestions.

Department of Chemistry,
University College,
Dublin.
W. P. Doyle.

29th May 1956.

Frerichs, R., 1953, *J. Opt. Soc. Amer.* **43**, 1153.
Gibson, A. F., 1950, *Proc. Phys. Soc.* B, **63**, 756.
Rekalova, G. I., 1951, *Zh. Tekh. Fiz.*, **21,** 1041, cf. *Chem. Abs.*, **46,** 2916.

CORRIGENDA

Dielectric Breakdown in Solids, by H. Fröhlich and B. V. Paranjape (*Proc. Phys. Soc.* B, 1956, **69,** 21).

(1) In equations (3.4 *a*, *b*) and (3.5 *a*), replace π by 2π. Six lines after equation (3.5 *b*), replace $q_0^2/2m$ by $q_0^2/8m$.

(2) Dr. R. Stratton has pointed out to us that in the development of the square brackets in the integral (2.25), leading to equations (3.4 *a*, *b*), the zero order term does not vanish completely (as assumed) because of the difference in the integration limits of $[\,]_+$ and $[\,]_-$. This leads to a factor $(1-4mkT/p^2)$ on the right-hand side of equations (3.4 *a*, *b*) case $2p<q_0$ and hence, six lines below equation (3.5 b), to the replacement of $2{\cdot}5kT_c$ by $1{\cdot}8kT_c$, and of $0{\cdot}9kT_c$ by $1{\cdot}6kT_c$. In equations (3.6 *a*) and (3.6 *b*), the factors 0·4 and 0·2 become 0·30 and 0·17 respectively and the right-hand sides of (4.3) must be multiplied by 3/4 and 17/20 respectively.

REVIEWS OF BOOKS

Chemistry of the Defect Solid State, by A. L. G. Rees. Pp. viii + 136. (London: Methuen; New York: John Wiley, 1954.) 8*s*. 6*d*.

This little book is not alone among little books in awaking a dubious attitude in the mind of the reviewer : the author has not yet shown that he can write a good big book about this subject, let alone a little one. A little book is ideally justified when the subject is so neatly circumscribed that it can be fully stated in a small compass, or when the author has a particular genius for laying bare the essential skeleton of a subject, or when he has another kind of particular genius for the encyclopaedic compression of much information into a small space. In the present case none of these conditions is satisfied, and a feeling of congestion results.

The author takes the opportunity of innovating in the language of the subject. The reviewer was surprisingly soon reconciled to the use of the word ' defect ' as an adjective as well as a noun : whether this is a reconciliation to be welcomed he is not sure. The author's descriptive notation is essentially that of chemical formulae and equations, with the added symbols □ and △ to represent the ' nonentities ', a (vacant) proper site and interstitial site of the structure.

As for the two other chief symbols in the notation, one defined as meaning ' distributed over ' is mostly used to mean ' combined with ', when it could advantageously be omitted. The other, meaning ' coexistent with ', could surely have been an ordinary +. A defect which this notation shares with others is the use of the affix + in the same position for totally different (and in one sense, opposite) meanings, as in Na^+, the sodium ion, and $\square^+$ the cation vacancy (which has an effective negative charge). The simplest remedy is to make the + a suffix in the latter case. Like most others in the subject, the author talks of ' lattice sites ' in the loose sense common among physicists and so annoying to the crystallographers. A reform in this matter is by no means easy, and physicists must quickly determine whether they will submit themselves to severe discipline, or declare firmly that the word ' lattice ' now has two different scientific meanings. The worst thing is to accept the situation unconsciously.

There is an odd claim on p. 49, that " it is possible to pass from one structure to another from which it is crystallographically distinct (even of different symmetry) without change of phase ". The argument leading to this conclusion has been published since the book, and does not satisfy the reviewer. Irrespective of its validity, it is surely wrong to attempt to be more than up to date in a book of this character.

Apart from the disregard of dislocations, the approach is thoroughly modern, and anyone researching or teaching in the field of chemical processes in the solid state, capable of exercising selection, is likely to find material of value to him in this book. Students will do better to get their information from here than not at all, but should be advised against using it as their sole source of information on the subject. Some of the diagrams are good. F.C.F.

CONTENTS OF SECTION A

PAGE

Dr. G. K. Woodgate and Dr. R. W. Hellwarth. Hyperfine Structure of Radioactive Silver ${}^{111}_{47}$Ag 581

Dr. G. K. Woodgate and Dr. R. W. Hellwarth. Shape of Resonance Curves in Atomic Beam Experiments 588

Mr. K. W. Taylor. α-Particle Counting Behaviour of Diamond Conduction Counters at Field Strengths up to 200 kv cm^{-1} 593

Dr. J. K. Perring and Dr. T. H. R. Skyrme. The Alpha-Particle and Shell Models of the Nucleus 600

Dr. A. Dalgarno, Mr. N. Lynn and Mr. E. J. A. Williams. The Oscillator Strengths of the 1sσ–2sσ, 1sσ–3dσ and 2sσ–3dσ Transitions of HeH^{2+} . 610

Dr. A. Dalgarno and Mr. M. R. C. McDowell. Charge Transfer and the Mobility of H^{-} Ions in Atomic Hydrogen 615

Mr. D. R. Chick, Mr. W. W. Evans, Mr. D. A. Hancock, Dr. S. E. Hunt and Mr. R. A. Pope. An Investigation of the γ-Rays from the Reactions ^{24}Mg(p, γ)^{25}Al and ^{12}C(p, γ)^{13}N, with a Note on the Positron Disintegration of ^{25}Al 624

Research Notes :

Dr. A. Dalgarno and Dr. J. T. Lewis. The Equivalence of Variational and Perturbation Calculations of Small Disturbances 628

Mr. H. N. V. Temperley. New Evidence about the Nature of the Liquid Helium Transition 630

Mr. C. A. Caine. The Magnetic Moments of Odd–Odd Nuclei . . 635

Dr. A. D. Buckingham, Prof. C. A. Coulson and Dr. J. T. Lewis. The Quadrupole Moment of a Hydrogen Atom in a Uniform Field . . 639

Dr. J. Varma. The Reaction ^{24}Mg(p, γ)^{25}Al. 641

Mr. A. E. Taylor and Mr. E. Wood. Proton–Proton Scattering at 134 mev 645

Letters to the Editor :

Miss M. K. Advani, Mr. P. Harihar and Dr. K. M. Gatha. Nuclear Charge Density Distributions 650

Contents of Section B 652

PROC. PHYS. SOC. VOL. 69, PT. 8—B (W. A. SMITH)

Figure 1. A polished surface.

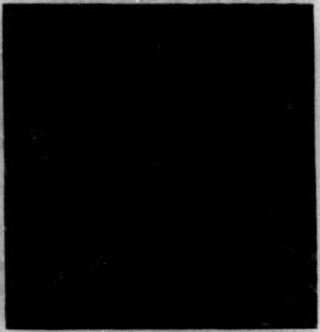


Figure 2. A surface after annealing, previously polished.

CONTENTS OF SECTION A

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

Letters to the Editor

[illegible]

Contents of Section B [illegible]

The Absorption of Ultrasonic Waves in a Number of Pure Liquids over the Frequency Range 100 to 200 Mc/s

By E. L. HEASELL† and J. LAMB

Electrical Engineering Department, Imperial College, London S.W.7

MS. received 12*th March* 1956, *and in final form* 18*th April* 1956

Abstract. The absorption of a sound wave in a liquid is conveniently represented by the quantity α/f^2 where α is the absorption coefficient relating to the excess pressure and f is the frequency. Values of α/f^2 are given in the paper for ninety-four liquids at a temperature of 25°c. The measurements were made by the pulse method at frequencies between 100 and 200 Mc/s and are estimated to be accurate to within $\pm 2\%$. The essential features of the apparatus are described.

The results are reviewed in the light of existing knowledge concerning the mechanisms responsible for attenuation of a sound wave in a liquid.

§ 1. Introduction

A study has been made in this laboratory of the mechanisms responsible for the absorption of ultrasonic waves in pure liquids. In this work, particular attention has been given to the investigation of anomalous absorption due to relaxation effects. At the same time considerable information has been accumulated concerning the absorption in liquids, which in many cases do not exhibit relaxation over the frequency range at present available.

Since, in the absence of relaxation, the absorption coefficient α relating to the excess pressure, is proportional to the square of the frequency f, it transpires that relatively small quantities of liquid—of the order of 10 cm^3—are required for absorption measurements in a suitably designed apparatus at frequencies above 100 Mc/s. This facilitates the study of a large number of liquids which would be prohibitive at lower frequencies where considerably larger quantities of liquid are generally required.

Measurements have been made by the pulse method over the frequency range 100 Mc/s to 200 Mc/s and at temperatures close to 25°c. The results are conveniently represented in terms of the quantity α/f^2 and values are given in this paper for ninety-four pure liquids. Since little reliable information is available concerning the absorption values in different liquids it is hoped that the present results, which are estimated to be accurate to within $\pm 2\%$ in α/f^2, will be of value to other workers in the field of ultrasonics.

The significance of the results is discussed in § 4 where the α/f^2 values in different liquids are compared in the light of existing knowledge concerning the mechanisms responsible for the absorption of sound waves in liquids.

---

† Now at the Research Laboratories, British Thomson-Houston Co. Ltd., Rugby.

§ 2. EXPERIMENT

2.1. *Experimental Method*

The absorption measurements have been made by the pulse method at selected frequencies between 100 and 200 Mc/s. In principle recordings are taken of the change in amplitude of a short pulse of sound waves as the path-length in the liquid is varied. A block diagram of the experimental system is given in figure 1.

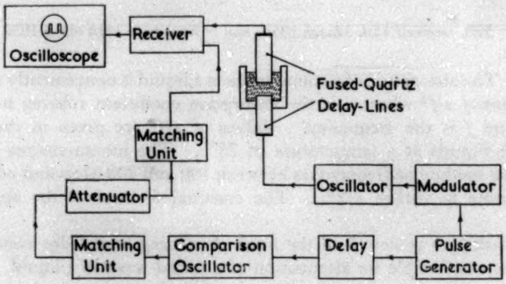


Figure 1. Schematic diagram of apparatus.

A high-power modulator produces pulses of up to 5 kv in amplitude and of 5 μsec duration at a repetition frequency of 500 per second. These pulses are used to modulate the anode of a self-excited oscillator which, in turn, produces bursts of oscillations of the required ultrasonic frequency. At the lowest ultrasonic frequency (100 Mc/s) each pulse contains 500 oscillations. The pulses of radio frequency are applied to an X-cut quartz crystal operating in the region of an odd harmonic of its fundamental mode of thickness vibration (20 Mc/s).

The transducer is acoustically coupled by means of a thin film of oil to one end of a fused-quartz delay line, both ends of which are accurately parallel and optically polished. Thus the acoustic pulse travels through the lower delay-line of figure 2, through the liquid under test and then into a second fused-quartz delay line, to the upper end of which is attached the receiving quartz crystal. The latter re-transforms the acoustic pulse into an electrical one which is then amplified, demodulated and finally displayed as a video pulse on the screen of the recording cathode-ray tube.

The vessel which contains the liquid under test is attached to the lower fused-quartz rod by means of a ground joint. The vessel has a double wall through which is circulated the temperature controlling liquid.

The change in amplitude of the first pulse received through the ultrasonic system is measured by employing a pulsed comparison oscillator operating at the same frequency as the main transmitter. This subsidiary oscillator produces short pulses of oscillations after a controllable time delay and the output is fed via a suitable matching network, to a piston attenuator operating in the E_{01} (circular) mode. The output from the attenuator is loosely coupled to the receiver by means of either a small loop or a probe placed near to the first tuned circuit of the receiver.

At the commencement of a measuring run the first received pulse from the ultrasonic system and the pulse from the comparison oscillator are arranged to be of equal amplitude and side by side on the screen of the oscilloscope. The liquid path is then reduced in known steps as registered on a calibrated dial gauge: at each setting of the path length the gain of the receiver is reduced by an appropriate amount and the attenuator is adjusted to equalize the pulse heights.

Attenuator readings, expressed in dB, are plotted against the path length in the liquid, the slope of the resulting straight line giving the absorption coefficient α of the liquid.

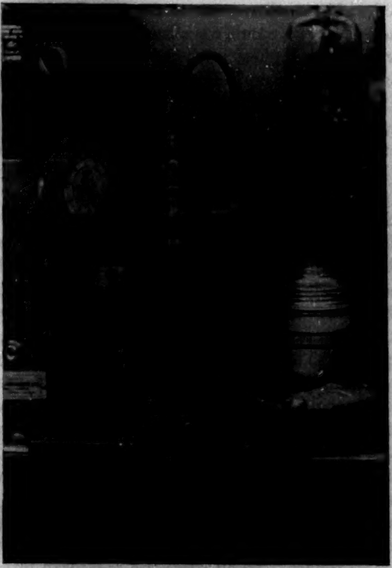


Figure 2. Photograph of the measuring apparatus (for description see text).

The frequency of the comparison oscillator is made equal to that of the main oscillator by causing the comparison pulse to overlap the first received pulse on the screen of the oscilloscope, and then adjusting the comparison oscillator for zero beat. Since there is no phase coherence between the two oscillators the zero beat condition is indicated by the appearance of a flat-topped region of varying height between the two pulses being received.

The frequency of oscillation is measured by means of a heterodyne wavemeter (Essen 1945) which is periodically checked against the Droitwich transmitter as a standard. Generally the frequency of the comparison oscillator is measured directly since, having a longer pulse length than the main oscillator, this produces a narrower frequency spectrum.

The temperature of the liquid is measured by means of a thermocouple mounted adjacent to the end of the movable fused-quartz rod: this ensures that the recorded temperature is that of the liquid in close proximity to the sound beam.

2.2. *Experimental Precautions and Accuracy*

All measurements were performed in pairs to ensure that the apparatus was yielding reproducible results and checks were made from day to day to protect against accidental contamination of the liquid under test. The containing vessel and the ends of the fused-quartz rods were cleaned after each liquid had been tested, and washed with a small amount of the next liquid to be measured prior to filling the vessel. Measurements were made on the liquids as supplied by the makers: these were of the highest purity available and where possible samples from different batches and from different suppliers were examined. Only in one case, that of ethyl alcohol, were the results found to vary appreciably from sample to sample, and this was attributed to the presence of water as an impurity.

In the results to be quoted in § 3 the major source of error was due to to thermal gradients in the liquid under test. The design of a containing vessel of small size, in order to permit the use of relatively small volumes of measuring liquid, precluded any stirring of the liquid. The temperature cycle of the thermostat was less than 0·05°c but it was found that the presence of very small thermal gradients in the liquid produced fluctuations in the height of the received pulse. This effect was particularly noticeable with liquids of comparatively high vapour pressure or with highly viscous liquids. Thermal stability was of greater importance in cases where a relatively long path length in the liquid (1 cm or more) was required. Thus the error from this cause was greater for liquids of low absorption towards the lower end of the frequency range.

Observation of the beating between the comparison oscillator and the main oscillator showed that their frequencies did not vary significantly over periods of half an hour, which was more than the time required for two measuring runs. Further the relative pulse amplitudes as checked by a peak voltmeter did not alter by more than ½% over the same time interval.

The E_{01} circular piston attenuator employed in this work was of accurate bore and was checked against other attenuators operating in the H_{11} (circular) and H_{01} (rectangular) modes. In all cases the measured attenuation constants confirmed the calculated values to better than ½%.

Less signal strength was available at the receiving quartz crystal at the higher frequencies so that in these cases the noise level of the receiver was a further source of error. However, in the present method of operation the receiver was only used at full gain when the first reading was taken corresponding to maximum path length in the liquid. The error due to noise was therefore much less than in other possible methods of arrangement—for example, with the attenuator placed in series with the acoustic path, in the absence of a comparison oscillator.

The reproducibility of successive measurements of α/f^2 taken at the same frequency on the same sample of liquid was generally of the order of ± ½%, but slight changes were observed between different frequencies. In the case of liquids which exhibit a consistent and reproducible decrease of α/f^2 with increasing frequency, this is attributed to the presence of a relaxing process.

Experimental Results

	Liquid	T_1	f_1	α_1/f_1^2	T_2	f_2	α_2/f_2^2	T_3	f_3	α_3/f_3^2
	Alcohols									
1.	Methyl alcohol (I)	24·1	104·1	30·1	—	—	—	24·9	192·4	30·6
2.	*n*-amyl alcohol (H)	24·8	104·9	97·0	24·9	148·0	97·2	24·8	192·0	96·5
3.	*iso*-amyl alcohol (L)	25·2	107·4	131	24·5	148·1	133	24·8	192·0	131
4.	*n*-butyl alcohol* (H)	24·8	105·0	81·1	25·0	148·0	79·1	25·1	192·0	81·4
5.	*iso*-butyl alcohol* (H)	25·0	107·4	153	25·0	148·1	146	25·1	192·1	140
6.	Ethylene glycol mono-butyl ether (H)	24·6	104·1	80·7	—	—	—	—	—	—
7.	Allyl alcohol (H)	24·8	104·1	44·1	—	—	—	24·8	192·0	45·4
8.	†3-5-5 trimethyl hexanol-1 (L)	25·2	107·4	294	25·7	148·1	269	25·5	192·2	263
9.	2-chloroethanol (B)	24·4	104·1	58·8	—	—	—	—	—	—
10.	Benzyl alcohol (H)	24·6	104·0	79·2	—	—	—	24·8	192·5	79·3
11.	Furfuryl alcohol (H)	24·4	104·0	90·5	—	—	—	25·0	192·0	91·8
	Carboxylic Acids									
12.	Acetic acid* (H)	24·7	104·0	153	—	—	—	24·8	192·0	139
13.	Dichloroacetic acid (H)	25·0	104·8	214	—	—	—	24·8	192·0	201
14.	Propionic acid (H)	24·4	104·1	131	—	—	—	—	—	—
15.	*iso*-valeric acid (H)	25·0	104·8	180	—	—	—	25·0	192·5	167
	Esters									
16.	Methyl acetate (H)	24·4	104·2	36·4	—	—	—	—	—	—
17.	Methyl formate (H)	24·1	104·0	49·3	—	—	—	—	—	—
18.	Ethyl formate (H)	24·1	104·0	49·9	—	—	—	—	—	—
19.	*n*-propyl acetate (B)	24·4	104·1	42·9	—	—	—	25·0	192·1	40·4
20.	*iso*-propyl acetate (B)	24·6	104·1	62·0	—	—	—	25·0	192·1	64·0
21.	*iso*-butyl acetate (H)	25·0	104·8	60·8	—	—	—	25·0	192·1	53·2
22.	Ethyl butyrate (B)	24·4	104·0	38·6	—	—	—	—	—	—
23.	Methyl butyrate (B)	24·5	104·1	35·3	—	—	—	—	—	—
24.	Dimethyl phthalate (B)	25·0	107·4	188	24·9	148·1	185	24·8	192·5	188
25.	Dibutyl phthalate (H)	25·0	107·0	250	25·0	148·1	244	24·8	192·7	237
26.	Methyl salicylate (H)	24·5	104·0	57·5	—	—	—	—	—	—
27.	Methyl chloroacetate (B)	24·4	104·0	42·7	—	—	—	—	—	—
	Amides									
28.	Formamide (B)	24·4	104·0	39·2	—	—	—	—	—	—
29.	*n*-methyl formamide (B)	24·6	104·1	32·6	—	—	—	—	—	—
	Amines									
30.	Diethylamine (H)	24·4	104·1	35·5	—	—	—	24·6	192·7	35·3
31.	Cyclohexylamine (H)	24·6	104·0	67·8	—	—	—	24·7	192·7	64·0
32.	Dicyclohexylamine (H)	24·4	104·0	157	—	—	—	24·6	192·7	146
33.	Ethanolamine (H)	24·8	104·0	166	—	—	—	24·8	192·7	159
34.	*o*-anisidine (H)	24·8	104·3	58·8	—	—	—	—	—	—
35.	*o*-chloroaniline (H, I)	24·4	104·0	56·2	—	—	—	24·8	192·1	53·7
	Aldehydes									
36.	Anisaldehyde (B)	24·6	104·1	63·6	—	—	—	—	—	—
37.	Cinnamaldehyde (H)	24·8	104·2	96·0	—	—	—	—	—	—
38.	Crotonaldehyde (H)	24·6	104·2	82·7	—	—	—	—	—	—
39.	*o*-chlorobenzaldehyde (B)	24·4	104·0	106	—	—	—	—	—	—
	Acetones									
40.	Acetone (H)	24·4	104·3	25·7	—	—	—	—	—	—
41.	Mesityl oxide (H)	24·4	104·1	35·7	—	—	—	—	—	—

† In order to confirm the downward trend of α/f^2 with increasing frequency an additional measurement was made on 3-5-5 trimethyl hexanol-1 at 15·2 Mc/s and 25°c: the α/f^2 value was 322×10^{-17} $\mathrm{sec}^2\,\mathrm{cm}^{-1}$.

	Liquid	T_1	f_1	α_1/f_1^2	T_2	f_2	α_2/f_2^2	T_3	f_3	α_3/f_3^2
	Ethers									
42.	Diethyl ether* (H)	24·4	104·1	44·5	—	—	—	—	—	—
43.	*iso*-propyl ether (H)	24·4	104·1	52·5	—	—	—	—	—	—
44.	Acetal (H)	24·6	104·1	56·7	—	—	—	—	—	—
45.	Methylal (H)	24·4	104·1	39·2	—	—	—	—	—	—
46.	Anisole (H)	24·8	104·0	43·8	—	—	—	24·6	192·1	43·5
47.	*o*-bromo-anisole (B)	24·4	104·0	71·3	—	—	—	24·6	192·1	68·5
48.	*p*-bromo-anisole (B)	24·3	104·0	63·6	—	—	—	24·6	192·1	61·1
49.	*o*-nitroanisole (B, I)	24·6	104·0	69·8	—	—	—	24·8	192·1	69·0
50.	Dioxan* (B)	24·6	104·1	117	—	—	—	24·8	192·7	112
	Benzene Derivatives and Related Compounds									
51.	Benzene* (M, H)	24·4	104·0	849	24·9	148·1	800	25·0	192·5	775
52.	Chlorobenzene (H)	24·4	104·0	147	—	—	—	24·6	192·3	146
53.	Bromobenzene (H)	24·8	104·0	144	—	—	—	—	—	—
54.	*o*-dichlorobenzene (H)	24·4	104·1	132	25·1	148·1	129	25·0	192·3	131
55.	*iso*-propyl benzene (H)	25·1	106·8	65·3	—	—	—	24·6	192·2	63·3
56.	*tert*-butyl benzene (B)	24·6	104·0	70·5	—	—	—	24·8	192·3	65·5
57.	Ethyl benzene (H)	24·4	104·1	56·0	—	—	—	24·8	192·3	54·3
58.	Benzyl chloride (H)	24·6	104·1	78·8	—	—	—	25·3	192·3	74·3
59.	Benzotrichloride (H)	24·4	104·0	110	—	—	—	24·8	192·3	105
60.	Nitrobenzene* (H)	24·4	104·0	74·2	25·0	148·1	72·2	24·8	192·3	74·1
61.	*β*-picoline (H)	24·4	104·1	66·1	—	—	—	24·8	192·6	64·0
62.	Toluene (H)	25·2	106·8	85·6	—	—	—	24·6	192·7	85·5
63.	*o*-chloro-toluene (B)	24·4	104·0	72·1	—	—	—	24·8	192·6	68·6
64.	*o*-xylene (H)	24·4	104·1	63·0	—	—	—	24·6	192·6	63·0
65.	*o*-chloro phenol (H)	24·4	104·0	97·6	—	—	—	25·0	192·2	93·9
66.	Indene (H)	24·6	104·2	133	—	—	—	24·7	192·4	132
67.	*p*-cymene (H)	24·4	104·1	52	—	—	—	—	—	—
	Cyclohexane and Derivatives									
68.	Cyclohexane (H)	25·0	107·5	192	25·0	148·0	190	24·8	192·3	182
69.	Methylcyclohexane (H)	25·0	107·0	98·3	24·8	148·0	93·8	24·8	192·3	92·9
70.	‡1-1 dimethylcyclohexane (A)	24·1	104·1	127	—	—	—	—	—	—
71.	‡*trans*-1-2-dimethylcyclohexane (A)	23·9	104·1	78·2	—	—	—	—	—	—
72.	‡*cis*-1-3-dimethylcyclohexane (A)	24·6	104·0	102	—	—	—	—	—	—
73.	‡*trans*-1-3-dimethylcyclohexane (A)	24·6	104·1	84·6	—	—	—	—	—	—
74.	‡*cis*-1-4-dimethylcyclohexane (A)	23·9	104·0	87·6	—	—	—	—	—	—
75.	‡*trans*-1-4-dimethylcyclohexane (A)	22·9	104·1	112	—	—	—	—	—	—
76.	Cyclohexanone (H)	24·6	104·0	72·8	—	—	—	24·8	192·0	70·7
77.	2-methylcyclohexanone (H)	24·6	104·1	106	—	—	—	25·0	192·1	100
78.	Cyclohexene (H)	25·0	107·0	103	25·0	148·1	102	25·5	191·8	101
79.	Deca-hydronaphthalene (B)	24·6	104·0	124	—	—	—	24·8	192·6	118
	Simple Halogen Derivatives									
80.	Bromoform (H)	25·1	106·9	262	25·3	148·1	256	25·2	192·2	248
81.	Chloroform (H)	24·4	103·5	363	—	—	—	25·2	192·2	365
82.	Carbontetrachloride* (H)	25·1	106·9	546	24·6	148·1	556	25·1	192·8	545
83.	Methylene chloride (H)	25·4	107·4	779	25·0	148·1	670	24·8	192·8	550
84.	Ethyl bromide (H)	25·0	106·9	78·7	24·8	148·1	78·4	24·6	192·7	75·5
85.	Ethylene bromide (H)	25·1	106·9	303	25·3	148·1	295	25·0	192·4	297
86.	Methyl iodide (M)	24·4	104·0	309	25·2	148·1	307	24·8	192·8	306
87.	1, 1, 1-trichloroethane (I)	25·1	106·8	436	25·0	148·1	434	24·8	192·6	431
	Miscellaneous									
88.	*n*-hexane (H)	25·0	107·5	60·4	25·0	148·0	58·5	24·9	191·5	56·7
89.	Epichlorohydrin (H)	24·4	104·0	68·1	—	—	—	—	—	—
90.	Acetyl chloride* (B)	24·6	104·0	82·8	—	—	—	—	—	—
91.	Acetic anhydride* (H)	24·6	104·0	57·7	—	—	—	—	—	—
92.	Water (distilled)	—	—	—	—	—	—	24·3	192·8	21·2
93.	Methyl disulphide (B)	24·9	104·0	58·7	—	—	—	—	—	—
94.	Carbon disulphide* (H)	24·8	104·0	2068	25·6	148·1	1210	25·5	189·2	776

α/f^2 in units of 10^{-17} sec^2 cm^{-1}, f (Mc/s), and T (°c).

H, Hopkin and Williams Ltd.; B, British Drug Houses Ltd.; I, Imperial Chemical Industries Ltd.; M, May and Baker Ltd.; A, American Petroleum Institute; L, L. Light and Co. Ltd. : * 'Analar' grade.

‡ Owing to the limited amounts of liquid available the estimated accuracy of α/f^2 is here ±5%.

Examination of the possible sources of error in conjunction with the consistency of α/f^2 determination at different frequencies for liquids which do not exhibit relaxation leads to an estimate that the values of α/f^2 are accurate to within $\pm 2\%$. As stated in the table, this accuracy was not achieved in the case of some dimethylcyclohexanes since insufficient quantities of these were available for measurement ($<5\,\text{cm}^3$).

§ 3. RESULTS

The experimental results are given in the table: in cases where measurements at more than one frequency are quoted, those at the highest frequency are more reliable since greater care was given to them.

The liquids are broadly grouped into classes of like compounds and each result is the average of at least two measurements, and generally of several.

§ 4. DISCUSSION

With the exception of water and ethyl alcohol (Pinkerton 1949) very little accurate information is available concerning the ultrasonic absorption in pure liquids. Only in the case of a few liquids measured by Rapuano (1947) have measurements been made at frequencies above 100 Mc/s. Increasing attention is now being given in several laboratories to the study of ultrasonic absorption. It is hoped that the data given in this paper will, to some extent, fulfil the need for standards of absorption in a number of liquids which are of reliable accuracy.

As a preliminary to comparing the absorption results in different liquids consideration will first be given to processes which are now known to govern the ultrasonic absorption in liquids. The measured absorption generally exceeds the classical contribution given by

$$\frac{\alpha}{f^2}=\frac{8\pi^2\eta}{3\rho c^3}$$

where η is the shear viscosity and ρ the density. The excess absorption is attributed to relaxation effects involving the perturbation by the sound wave of a molecular equilibrium. If τ is the relaxation time associated with the establishment of equilibrium following a sudden distrubance to the system, then in the case of perturbation by a sound wave the absorption per wavelength μ $(=\alpha c/f)$ attains a maximum in the neighbourhood of the characteristic frequency $f_c(=1/2\pi\tau)$. The value of α/f^2 decreases with increasing frequency from its low-frequency value ($f \ll f_c$) to a limiting high-frequency value ($f \gg f_c$).

The equilibrium which is perturbed may be of an intra- or an inter-molecular nature. An example of relaxation arising from an intra-molecular equilibrium is afforded by carbon disulphide (No. 90) where it has been shown (Andreae, Heasell and Lamb 1956) that the anomalous absorption is due to a relaxation of the total vibrational specific heat. Alternatively a molecule may be able to assume two or more configurations which differ in energy, as a result of rotation about certain bonds.

A monomolecular reaction of this type has been shown to be responsible for the anomalous absorption observed in a number of cyclohexane derivatives (Karpovich 1954, Lamb and Sherwood 1955) and in a range of unsaturated aldehydes (de Groot and Lamb 1955, 1956). The latter authors have also argued (1955) that the relaxation effects observed in esters can be attributed to the existence of two planar configurations stabilized by resonance.

Excess absorption arising from an equilibrium of an inter-molecular nature has been proposed for water by Hall (1948) but in this case the characteristic frequency is clearly much higher than the experimental range now available. In this instance it was proposed that the equilibrium is established between an open structure of the molecules and a more tightly packed or solid-like structure. It is generally supposed that a similar mechanism is responsible for the excess absorption in alcohols. In this connection it is noteworthy that the α/f^2 value for water (No. 88) measured at 192·8 Mc/s and 24·3°c is in exact agreement with the value obtained by Pinkerton (1949) over the frequency range 7·5 to 67·5 Mc/s.

Considering the results of § 3 in the light of the above it can be seen that for most of the alcohols α/f^2 is independent of frequency and generally increases with increasing chain length. The exceptions are *iso*-butyl alcohol (No. 5) and 3–5–5 trimethyl hexanol-1 (No. 8). Here there is definite evidence of relaxation which may be of an inter-molecular structural type and thus merits further investigation. On the other hand the consistent decrease of α/f^2 with increasing frequency may be due to the presence of small amounts of water as an impurity. The results of Storey (1952) indicate that the characteristic frequency of certain ethyl alcohol–water mixtures is in the region of 100 Mc/s. In this preliminary investigation of a large number of liquids no attempt was made to purify these compounds further.

The carboxylic acids are known to exhibit relaxation: thus for acetic acid (No. 12) $f_c = 0{\cdot}725$ Mc/s at 25° c (Lamb and Pinkerton 1949) whilst for propionic acid (No. 14) $f_c = 2{\cdot}02$ Mc/s at 21°c (Lamb and Huddart 1950). In both cases the present measurements lie on the 'high-frequency tail' of the relaxation and invalidate the previous suggestion of a second relaxation in acetic acid at frequencies not much above 100 Mc/s. This was put forward by Lamb and Pinkerton to explain their results at the highest frequency of measurement then available (67·5 Mc/s). Clearly the previous results at that frequency were in error.

Relaxation in esters has been observed by a number of experimenters, notably Biquard (1936), Libermann (1949), Huddart (1950), Pinkerton (1951) and Karpovich (1954). A probable explanation for these effects has been put forward by de Groot and Lamb (1955). At frequencies of 100 Mc/s and above the α/f^2 value in the short chain esters has almost reached its limiting high-frequency value but in some of the longer chain compounds, for example in *iso*-butyl acetate (No. 21), α/f^2 is still decreasing slightly with increasing frequency.

The unsaturated aldehydes form a group in which relaxation occurs (de Groot and Lamb 1955, 1956) due to the existence of two alternative planar configurations stabilized by resonance.

With regard to the other liquids of the table considerable interest centres round the results for benzene. There is clear evidence of a relaxation in this liquid but the characteristic frequency is above the present available range of measurement. Presumably the effect is associated with the internal vibrational specific heat of the benzene molecule but a theoretical analysis cannot be made until results at higher frequencies are available. Possibly the decrease in α/f^2 with increasing frequency found in a number of benzene derivatives may be associated with the same phenomenon although in some cases, for example *o*-chlorobenzaldehyde (No. 39) there exists the possibility of alternative configurations of the groups attached to the benzene ring.

Finally, the relaxation in methylene chloride (No. 83) is also, presumably, associated with the vibrational specific heat. Relaxation in this compound was first observed in this laboratory by Andreae (1955) and the present results are an extension of his work. Too much reliance must not be placed in the numerical values for this liquid since it forms a constant boiling mixture with water and additional purification is required. This is at present receiving attention from Dr. Andreae.

ACKNOWLEDGMENTS

The work described in this paper was supported by a grant from the Department of Scientific and Industrial Research. The authors gratefully acknowledge this assistance and E.L.H. wishes also to thank the same body for provision of a maintenance grant during the course of this work. The authors wish to thank Dr. J. H. Andreae for helpful co-operation and discussion, and the Esso Development Co. Ltd. who kindly provided the dimethylcyclohexanes.

REFERENCES

ANDREAE, J. H., 1955, *Ph.D. Thesis*, University of London.
ANDREAE, J. H., HEASELL, E. L., and LAMB, J., 1956, *Proc. Phys. Soc.* B, **69**, 625.
BIQUARD, P., 1936, *Ann. Phys., Lpz.*, **6**, 195.
ESSEN, L., 1945, *J. Inst. Elect. Engrs*, Pt. III, **92**, 291.
DE GROOT, M. S., and LAMB, J., 1955, *Trans. Faraday Soc.*, **51**, 1676; 1956, *Nature, Lond.*, **177**, 1231.
HALL, L., 1948, *Phys. Rev.*, **73**, 775.
HUDDART, D. H. A., 1950, *M.Sc. Thesis*, University of London.
KARPOVICH, J., 1954, *J. Chem. Phys.*, **22**, 1767.
LAMB, J., and HUDDART, D. H. A., 1950, *Trans. Faraday Soc.*, **46**, 540.
LAMB, J., and PINKERTON, J. M. M., 1949, *Proc. Roy. Soc.* A, **199**, 114.
LAMB, J., and SHERWOOD, J., 1955, *Trans. Faraday Soc.*, **51**, 1674.
LIEBERMANN, L. N., 1949, *Phys. Rev.*, **75**, 1415; **76**, 440.
PINKERTON, J. M. M., 1949, *Proc. Phys. Soc.* B, **62**, 129; 1951, *Ultrasonics Conference, Brussels*, p. 117 (*Med. Vlaamsche Akad. Kl. Wet.*).
RAPUANO, R. A., 1947, *Phys. Rev.*, **72**, 78.
STOREY, L. R. O., 1952, *Proc. Phys. Soc.* B, **65**, 943.

An X-ray Study of the Structure of a Sodium Chloride Crystal Grown by the Kyropoulos Method

By D. A. JONES and T. SMITH
Department of Natural Philosophy, Marischal College, Aberdeen

Communicated by R. V. Jones; *MS. received* 18*th July* 1955,
and in revised form 22*nd March* 1956

Abstract. An X-ray examination has been made of the crystallite structure of a sodium chloride crystal grown with a varying growth velocity by the Kyropoulos method. The crystallites are axial in form and their major misorientation corresponds to a rotation about the boule axis and varies little with growth rates between 0·04 and 0·6 cm hr^{-1}, though the cross-sectional area decreases. The structures found at the high growth rates cannot be completely explained by the mechanism proposed by Teghtsoonian and Chalmers, as evidence of polygonized regions was obtained.

---

§ 1. Introduction

Single crystals of salt grown by the Kyropoulos method contain many small regions of slightly different orientations. The presence of these crystallites may be inferred from the appearance of a cleavage surface which consists of the cleavage planes of the separate crystallites. The study of the angles between the planes does not, however, form a very convenient method of investigating the structure of the underlying crystal as the cleavage surface is complicated by tear markings which are characteristic of the mode of propagation of the rupture stress rather than the crystal structure (Zapffe and Worden 1949). There is also the possibility that cleavage takes place through regions where the crystallites have well aligned cleavage planes, so that studies from the cleavage surfaces may under-estimate the average misorientation between crystallites.

For these reasons it was decided to examine the material along the axis of the boule by the X-ray method first used by Guinier and Tennevin (1949). In this way the spread of diffracting material from fairly large specimens taken from different points along the boule axis was quickly examined with considerable sensitivity and resolution.

§ 2. Preparation of Specimens

The specimen boule examined had its growth direction along a cubic axis, and was typical of crystals grown by the Kyropoulos method. Its maximum diameter and height were 19 cm and 17 cm respectively.

The technique for growing these crystals was first of all to adjust the temperature of the furnace so that the crystal which formed on the seed grew at a steady rate. In the initial stages the rate of growth in the vertical direction was about 2 cm per hour. The furnace was held at this temperature while the rate of growth decreased to less than 0·04 cm hr^{-1} when the depth of crystal was 4–5 cm below the bottom of the seed. The temperature of the furnace was then decreased so that the crystal began to grow again with a velocity of about

0·6 cm hr^{-1}. This cycle was repeated until as much as possible of the melt had solidified on the crystal.

When growth was completed the crystal was removed to an annealing furnace and cooled over a period of three days.

The top 11 cm of the boule was of good optical quality, and it was material from along the axis of this part which was examined. The exterior of the lower part of the crystal showed arborescent growth and the interior was slightly milky, both effects of the increasing impurity contant (see figure 1, Plate).

The specimens to be examined were required in the form of plates 1–3 mm thick and with an area of about 1 cm square, and considerable care was devoted to preparing suitable specimens without introducing further strains. An axial column was extracted from the boule by cleavage and then reduced to the final specimen plates by splitting off very thin sheets. Under these conditions it is probable (Pratt 1953) that most of the strain introduced will be in the sheets that are discarded. Any strain that is produced will be concentrated near the surface of the specimen as it is there that the tension and compression resulting from cleavage is most intense.

In fact, for the plates used, the type of strain which can be investigated by the Guinier–Tennevin method shows no considerable increase near the surfaces of the specimen sheets. This was shown when similar results were obtained on x-ray examination before and after the specimen thickness had been reduced to half its original value by etching in water.

§ 3. EXPERIMENTAL METHOD

If white x-rays diverge from a point source to fall, at a distance D, on a specimen plate set up vertically with the normal to its surface making an angle α with the incident beam, then the radiation diffracted from a vertical set of crystal planes, the normals to which make an angle ϕ with the surface of the specimen, will be brought to a vertical line focus at a distance D' which is easily calculated. For the important case where α and ϕ are small, that is when the diffracted beam suffers little deviation, then $D' = D$ and the length of the trace produced is about twice the height of the specimen.

If the specimen consists of regions with two different orientations then two vertical traces will be produced, the separation between them being a measure of the rotation about a vertical axis. The vertical movement of the traces is insensitive to rotation about the horizontal axis.

In most of these experiments a horizontal strip of specimen 3·0 mm deep and 1 cm long was exposed, though for some of the exposures a square was irradiated. The resulting diffraction pattern from the {100} planes consisted of a number of short vertical traces. However, the rotation of the reflecting planes about a horizontal axis was sufficient to make it impossible to photometer all the traces along a line perpendicular to their length. The length of the traces was therefore increased by replacing the spot source of x-rays by a vertical line source, the foreshortened width of which was 0·35 mm. For a perfect specimen the width of the trace produced would also be 0·35 mm (neglecting the very slight increase due to the thickness of the specimen plate). With a value of D' of 70 cm it was found possible to resolve traces separated by 0·3 mm which correspond to a spread in the normals of the reflecting planes of 0·75′.

Exposures of 2 hr at 20 mA and 40 kv were required. The traces were photometered on a Hilger recording microphotometer, and conversion of the results into intensity curves was facilitated by recording calibration strips on the films by exposing them behind a rotating sector.

The integrated intensity from each crystallite was proportional to the area under the corresponding line on the final intensity curve, and will be proportional to the volume of the crystallite if, as is the case for most NaCl crystals, the primary extinction is unimportant.

§ 4. RESULTS

Before describing the detailed relationships between the crystallites the character of their orientation range measured in different directions will be given in terms of the half spread, that is the wedge angle within which is included the normal to half the total volume of the specimen. This is used for a rough description as it gives an appropriate measure of the crystal perfection for many macroscopic purposes.

The half widths have been obtained for the cube axes in the different cube planes for specimens taken from the boule axis at different distances below the seed. If the cube axis parallel to the growth axis is [001] (figure 1), the first set of specimens were cleaved in the (001) plane and from them the average half spread of [100] in (010) was found to be 3·0′. From a similar set of plates the half spread of [010] in (100) was 2·7′. Although there is some fluctuation in the half spreads from different specimens (2·4′ to 4·0′ for [100] in (010)), there is no systematic variation to be observed with growth velocity.

The number of crystallites in the irradiated volume will be approximately equal to the number of traces on the photographs. Crystallites counted in this way will be those greater in volume than 0·04 mm^3, which is the estimated volume of the smallest crystallite counted and will be taken as the lower limit of observation; smaller crystallites presumably account for the unresolved background which is never observed to be more than 5% of the total reflected intensity, and in most cases is less than half this. It must also be noted that regions will be counted as separate crystallites only if separated by about 0·5′ which will show as a combined trace of observably increased width. Such broadened lines were seldom observed, the width of the traces being usually equal to that expected from the dimensions of the line source. The estimate of the number of crystallites defined in this way has been obtained by correcting for the possibility that two crystallites are counted together due to a chance coincidence between their reflecting planes. This probability can be obtained from the area taken up by the traces within the area through which the traces are distributed. For the photographs showing the spread of [100] in (010) and [010] in (100) this was estimated to be 1 in 12, so the number of crystallites was obtained by multiplying the number of traces by 12/11.

It was found that the number of crystallites estimated in this way showed no change as the plate thickness was increased from 1 mm to 3·0 mm, which was the limit set by the broadening of the traces. It was concluded, therefore, that the crystallites ran through the plates, their lengths along the boule axis being not less than 3·0 mm. The average cross-sectional area of these regions in the (001) plane shows considerable variation. Over the section of the boule axis where the rate of growth was decreasing the cross-sectional area of the crystallites increases steadily. The sharp fall in cross-sectional area between 4 and 5 cm (figure 2) corresponds to the increase in vertical growth rate in this region from 0·04 cm hr^{-1} to 0·6 cm hr^{-1}.

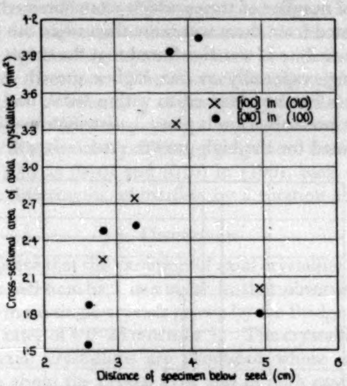


Figure 2. Cross-sectional area of crystallites at different points along the boule axis; observations from photographs showing variation of [100] in (010) and [010] in (100).

It is not possible to obtain information on the shape of the crystallites in the (001) plane if they are arranged irregularly, but certain other arrangements would be detected. Such an arrangement would consist of crystallites having a fusiform cross section disposed in a regular way; for instance, if all the long axes were parallel to one another, or all radially directed. This type of arrangement is excluded by the observation that the same cross section was obtained when the irradiated strip 3 mm × 1 cm was oriented in different ways in the (001) plane. A regular array of crystallites with cross-sectional axial ratio of 3 : 1 or over would have resulted in significantly different cross-sectional areas being obtained from the different photographs.

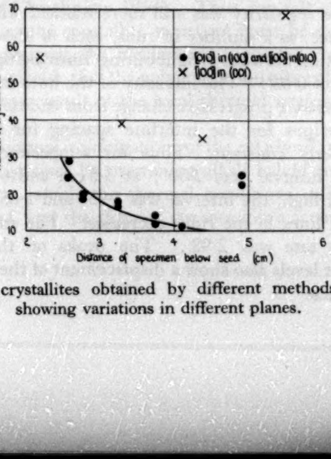


Figure 3. Number of crystallites obtained by different methods: from photographs showing variations in different planes.

Information on the variation of [100] in (001) was obtained from a separate set of specimens cleaved parallel to (100). When similar volumes were examined the half-spread had an average value of 15·4′, the maximum being 19·5′ in the high growth rate regions. However, the most noticeable feature of the photographs was the large number of traces which were obtained, so that the number of crystallites estimated from them was more than twice the previous estimate for all growth rates (figure 3). It was also found that the traces showed considerable regularity in spacing, especially at the higher growth rates. The intervals between trace centres can be obtained to within 0·05′, and when the frequency with which different intervals occur is plotted, distributions similar to those shown in figure 4 are obtained for the high growth rates. It will be seen that there is a sharp maximum for an interval corresponding to a separation between the corresponding planes of 1·75′. The decrease in the number of intervals below this is not due to any lack of resolution, for if such intervals were not resolved, the width of the lines would have been noticeably increased. The distribution in figure 4 can therefore be taken as giving the frequency with which different angles occur between crystallites of above 0·04 mm^3 when they are arranged in angular order. The regularity was still more marked when the intervals were examined in sequence, as a number of runs, each of about ten lines, was then observed, the intervals between neighbouring lines being constant within the experimental limits of 0·05′. The intensity of the lines in such a run is usually either constant or shows a progressive change from end to end. For the highest growth rates the values for the interline spacing for the runs on different photographs are fairly constant. Thus for a specimen taken from 1·1 cm below the seed the interval was 2·00′; at 5·5 cm and 6·6 cm, where the rate of growth was again high, the interval was 1·95′ and 1·85′. For lower growth rates the number of lines in the runs decreases. The maximum value found with a low growth rate was 2·98′. The peaks of the frequency interval curves from different levels also show a displacement of the peak to higher values for lower growth rates.

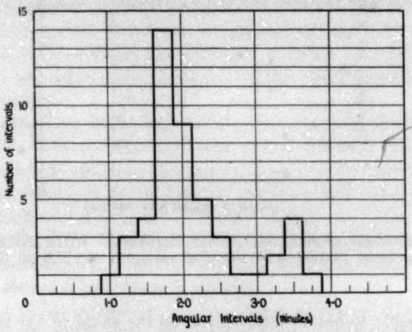


Figure 4. Frequency of different angular intervals between crystallites corresponding to adjacent traces on high growth rate photographs.

The shape of the angular frequency curve and the occurrence of the runs suggested that the angles between neighbouring crystallites might not be random for higher growth rates. It was found that not only the runs, but also the longer sequences within which the runs were often found, could be removed from one end as the length of the irradiated strip was reduced (figure 5, Plate).

Thus a feature of the higher growth rate material is the occurrence of regions in which groups of neighbouring crystallites have boundaries, the tilt in each of these being in the same direction. The runs are simply neighbouring crystallites for which the variation in boundary angle is very small indeed. The whole sequence of related crystallites may have a combined volume of 60 mm^3 and extend through many of the axial crystallites giving rise to the traces corresponding to the variation of [100] in (010) and [010] in (100), each of these crystallites consisting of regions differing in orientation by a rotation about the boule axis.

§ 5. DISCUSSION

These results indicate that the variation of axial crystallite size and orientation on growth rates below 0·6 $\mathrm{cm\,hr}^{-1}$ is similar to that observed by Teghtsoonian and Chalmers (1952) in tin single crystals grown by the Bridgman method with the much higher growth rates of 4·0–28 $\mathrm{mm\,min}^{-1}$. The crystallite size is much less, but in both cases axial crystallites are produced whose major disorientation consists of a rotation about the crystal axis, and in both cases the crystallite size varies inversely with the growth rate. The angular misorientations are also very much smaller than for tin (several minutes instead of several degrees), but in both materials the misorientation increases with growth rate. This similarity suggests that the mechanism of vacancy condensation and subsequent dislocation formation suggested by Teghtsoonian and Chalmers for tin may also be applicable to salt crystals grown under the above conditions.

For growth rates in the neighbourhood of 0·6 $\mathrm{cm\,hr}^{-1}$ a further structure is imposed on the axial crystallites which is not observed in tin even at much greater growth rates, and which cannot be readily explained by the Teghtsoonian and Chalmers mechanism. Structures of this kind are found, however, in many crystals which have been plastically bent and subsequently annealed. The length of the crystallite chains and the angles between their ends show that curvatures with radii between 18 and 25 cm would have been required, and rock salt crystals with these curvatures would be expected to show polygonization when annealed at a high temperature (Amelinckx 1954). Thus for salt crystals grown at all but the lowest velocities the crystallite structure must be explained in terms of polygonization to supplement the mechanism suggested by Teghtsoonian and Chalmers. It seems probable that the polygonization takes place near the growing surface of the crystal as a result of the local thermal strains associated with the high growth rates.

It is proposed to continue investigations of the relationship between crystallite structure and thermal stresses by examining alkali halide crystals grown by the Stockbarger method in which the thermal conditions during growth can be more readily controlled, and the thermal stresses more accurately estimated.

ACKNOWLEDGMENTS

We wish to express thanks to Professor R. V. Jones for his interest and encouragement, to Professor F. C. Frank for useful discussions and to Mr. G. Ingram who assisted in growing the crystal.

REFERENCES

AMELINCKX, S., 1954, *Acta Met.*, **2**, 848.
GUINIER, A., and TENNEVIN, J., 1949, *Acta Cryst., Camb.*, **2**, 133.
PRATT, P. L., 1953, *Acta Met.*, **1**, 692.
TEGHTSOONIAN, E., and CHALMERS, B., 1952, *Canad. J. Phys.*, **30**, 388.
ZAPFFE, C. A., and WORDEN, C. O., 1949, *Acta Cryst., Camb.*, **2**, 383.

The Determination of Dynamic Moduli and Internal Friction of High Polymers from Creep Measurements

By J. J. BENBOW

Akers Research Laboratories, Imperial Chemical Industries, Ltd., Welwyn, Herts.

MS. received 11*th January* 1956

Abstract. An account is given of the use of Fourier analysis to transform creep measurements to dynamic quantities with specific reference to some measurements made on polythene. The method is based on a technique described by Roesler, but additional theory is given which eliminates the necessity of determining the retardation spectrum explicitly. Using the derived dynamic quantities as a standard it is shown that more approximate methods provide sufficiently accurate transformations for many practical purposes if the retardation spectrum of the material is very flat.

---

§ 1. INTRODUCTION

In both fundamental research and technological problems concerned with visco-elastic materials it is desirable to know the elastic moduli and internal friction explicitly over wide frequency or time ranges. This may be achieved by making dynamic measurements (i.e. measurements involving sinusoidal stresses and strains) in which results are obtained at discrete frequencies. Many experiments are, however, required to cover an appreciable frequency range and at frequencies below one cycle per minute the experiments are difficult to perform and take much time. On the other hand, at long times, creep measurements (i.e. the study of the strain, caused by a constant stress, as a function of time) are relatively much easier and quicker to do. For instance, by applying the stress for one day the creep behaviour at all times from about 1/10 second to one day can be measured.

Creep experiments do not provide a direct measurement of the elastic moduli and internal friction of a material although in theory they contain the same information as dynamic measurements in a different form. The transformation of creep curves to dynamic quantities is difficult since it depends in principle on the determination of the retardation spectrum. (The retardation spectrum is a distribution of compliance over a time scale: see §3 and the references therein for detailed definition.) Several approximate methods have been put forward for doing this. The first of these was proposed by Alfrey (1948) and this involved dividing the creep curve by its differential with respect to ln (time). Later, Schwarzl and Staverman (1952) suggested that more accurate results could be obtained by using higher derivatives in addition to the first. An account of, and comparison between, these and other approximate methods is given by Leaderman (1954). Recently, Roesler and Pearson (1954), Roesler (1955) and Roesler and Twyman (1955) have described systematic procedures for determining relaxation and retardation spectra from all types of measurements on visco-elastic substances. These computational methods can also be used to transform results from one type of experiment into those obtainable from others, and it is for this application that they are likely to be of the greatest practical importance. By analysing experimental creep curves into Fourier series the transformations to dynamic quantities are readily performed and the accuracy is limited only by that of the experimental results. In this report an account is given of the application of Roesler's method to creep measurements made on polythene and the results are compared with those given by applying Alfrey's method to the same creep curve and with some direct dynamic measurements on polythene.

The comparisons show that dynamic properties are readily derived from creep curves and that, as expected, in the case of polythene which has a very flat retardation spectrum the approximate methods give results which differ very little from true values.

§ 2. EXPERIMENTAL METHOD

Before the transformation from creep to dynamic results can be made it is necessary to know the creep behaviour over a very wide time scale. In fact theoretically the creep curve should be known over a domain wide enough for it to become time independent at both extremes. In practice accurate measurements over six decades of time are sufficient to produce accurate dynamic quantities over five decades. The extrapolations used to make the curve tend to constant values at each end of the observed domain are likely to produce errors for about half a decade inside the observed range, so these results are discarded after the transformations have been made. (With many high polymers the large time range may be covered by using the Ferry (1950) reduction technique whereby results at different temperatures over limited time ranges are combined to give a single creep curve but over an extended time scale. However, this reduction was not used since it was found that six decades of time could be readily covered by direct measurements at one temperature.)

One of the most direct ways of applying a constant torsional stress to a mechanical system is by the use of weights and pulleys, but the shortest time after which measurements may be made is limited by the inertia of the weights and the difficulty of applying the stress suddenly. To cover a large number of decades it was necessary to extend the readings for a very long time. For example, if the first reading was made 10 seconds after the stress was applied, to cover six decades of time the last reading would have to be taken 120 days later. By using an electrical method of applying the stress so that readings were started after 0·04 second six decades were covered 12 hours later.

In the apparatus, which is shown diagrammatically in figure 1, the sample rod of polythene was fixed rigidly at its lower end and carried a coil of tinned copper wire (40 turns of No. 30 s.w.g.) at its upper end. A large permanent magnet was mounted so that its field was parallel to the plane of the coil. Thus a d.c. current passed through the coil caused a torsional stress to be applied to the polythene sample, the initial application being very rapid since it was only limited by the build up of the magnetic field after switching on the current. To calibrate the transducer the strain produced in a standard torsion wire by a known current was determined and it was found that within the five degrees of arc through which the coil rotated during the creep experiments, the torque was independent of

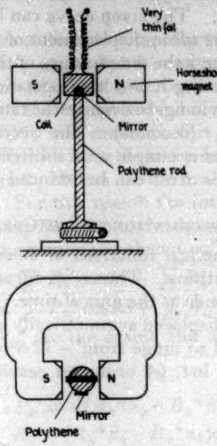


Figure 1. Creep apparatus.

the angle of twist. The lead-in wires to the coil were made of very flexible copper foil since it is essential to ensure that extraneous torques are reduced to a minimum. To avoid errors due to static friction no bearings were used, the sample being sufficiently robust to support both its own weight and that of the coil. A mirror of focal length one metre was attached to the upper end of the sample to measure the strain. In the interval between 0·04 second and 100 seconds a recording camera measured the rotation of the sample and from 10 seconds onwards visual readings were made on a graduated scale. . For convenience the readings from both the record and the scale were made at intervals which were, roughly, equally spaced when plotted on a logarithmic scale. Immediately after the stress was applied transient oscillations were set up which decayed rapidly due to the high

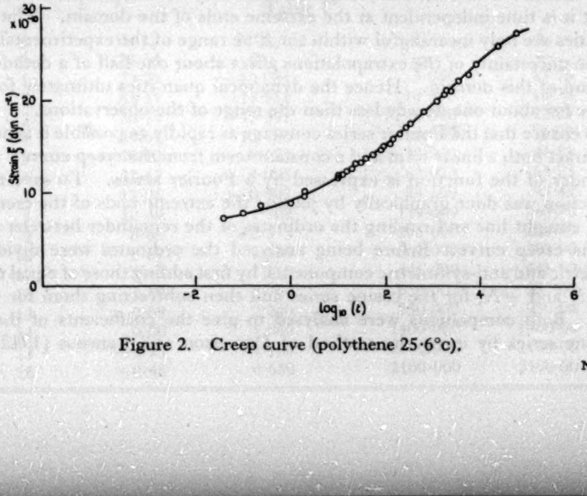


Figure 2. Creep curve (polythene 25·6°c).

internal friction of polythene. The creep curve can be extended into part of the transient region by finding the mean displacement of the oscillatory deformation.

Throughout the experiments the temperature of the sample was held constant to within ±0·02°c by immersing it in a thermostatically controlled water bath. A typical creep curve is shown in figure 2 where the ratio of strain to stress, denoted by ξ, is plotted against $\log_{10} t$ (seconds). The overall error to which repeated creep measurements on the same sample were subjected was about 5%; by using a larger recording camera this error can be reduced considerably.

§ 3. Transformation of the Creep Curve

In order to use the numerical functions of Roesler's paper, the time scale has to be in Napierian logarithms. The creep curve is shown replotted in this way in figure 3 with 100 seconds as the unit of time. This unit is chosen so that the experimental points are displaced symmetrically about the origin on the time axis. This scale is extended to range from $-12{\cdot}8$ to $+12{\cdot}8$ so that when subdivided into steps of 0·4 in $\ln t$, 64 ordinates result. Since the experimental points extend over a finite time range it is necessary to extrapolate the creep curve so that it is time independent at the extreme ends of the domain. The derived quantities are only meaningful within the time range of the experimental results, and the uncertainty of the extrapolations affect about one half of a decade within each end of this domain. Hence the dynamical quantities ultimately found are reliable for about one decade less than the range of the observations.

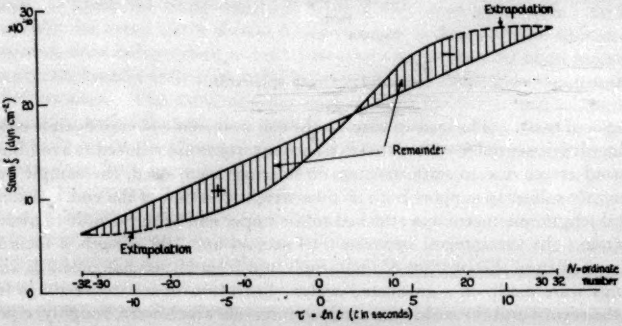


Figure 3. Analysed curve.

To ensure that the Fourier series converge as rapidly as possible it is necessary to subtract both a linear term and a constant term from the creep curve; only the remainder of the function is expressed by a Fourier series. To save time the subtraction was done graphically by joining the extreme ends of the creep curve with a straight line and reading the ordinates of the remainder between this line and the creep curve. Before being analysed the ordinates were divided into symmetric and anti-symmetric components, by first adding those of equal number, i.e. $+N$ and $-N$, for the cosine series and then subtracting them for the sine series. Both components were analysed to give the coefficients of the cosine and sine series by using the method of Danielson and Lanczos (1942). The coefficients of both components decreased rapidly with ascending order and those higher than the ninth clearly represented random errors and were neglected (see table 2). From these cosine and sine coefficients of the creep curve the corresponding coefficients of the complex dynamic compliance were computed.

In Roesler's paper (1955) the methods for determining the retardation spectrum or dynamic compliances from the creep function are given in principle, but only in the first-mentioned problem are the necessary formulae and numerical functions given explicitly. For this reason the formulae and other data for the direct transition from the creep function to the dynamic compliances will be given below. If J is the complex compliance, then $J = J' + iJ''$ and

$$J' = \sum P_k' \cos kx + \sum R_k' \sin kx \qquad \ldots\ldots(1)$$

$$J'' = \sum P_k'' \cos kx + \sum R_k'' \sin kx \qquad \ldots\ldots(2)$$

where P_k', R_k', P_k'' and R_k'' are the cosine and sine components of J' and J'' respectively.

It can be shown that

$$P_k' = -(A_k^* \alpha_k + B_k^* \beta_k)\phi_k \qquad \ldots\ldots(3)$$

$$R_k' = (A_k^* \beta_k - B_k^* \alpha_k)\phi_k \qquad \ldots\ldots(4)$$

$$P_k'' = (A_k^* \beta_k - B_k^* \alpha_k)\psi_k \qquad \ldots\ldots(5)$$

$$R_k'' = (A_k^* \alpha_k + B_k^* \beta_k)\psi_k \qquad \ldots\ldots(6)$$

where

$$\phi_k = \frac{\pi k}{L^2(\alpha_k^2 + \beta_k^2)} \left[\frac{2}{\pi} \sinh\left(\frac{k\pi^2}{2L} \right) \right]^{-1} \qquad \ldots\ldots(7)$$

$$\psi_k = \frac{\pi k}{L^2(\alpha_k^2 + \beta_k^2)} \left[\frac{2}{\pi} \cosh\left(\frac{k\pi^2}{2L} \right) \right]^{-1} \qquad \ldots\ldots(8)$$

A_k^* and B_k^* are the Fourier coefficients of the creep curve, α_k and β_k are the Fourier coefficients of the kernel, $L = 12{\cdot}8$, i.e. the half width of the domain of $\ln t$, and k is the order of the coefficients.

Since the functions α_k, β_k, ψ_k and ϕ_k are independent of experimental results it is convenient to tabulate them below for use with any creep curve. Their values for k ranging from 0 to 16 are given in table 1.

Table 1

k	$10^3\alpha_k$	$10^3\beta_k$	ψ_k	ϕ_k
0	78·14	0	—	—
1	73·75	−10·084	5·056	13·76
2	63·11	−15·511	10·870	16·78
3	50·77	−15·591	18·340	22·37
4	39·67	−12·408	28·530	31·27
5	30·52	− 8·123	43·030	44·88
6	23·17	− 4·095	63·970	65·23
7	17·26	− 0·903	94·560	95·42
8	12·49	1·300	139·600	140·20
9	8·68	2·375	205·600	206·00
10	5·69	3·120	302·000	302·90
11	3·43	3·110	445·000	445·90
12	1·80	2·760	650·000	650·00
13	0·69	2·240	950·000	950·00
14	0·00	1·640	1400·000	1400·00
15	−0·35	1·100	2100·000	2100·00
16	−0·48	0·650	3100·000	3100·00

With these values and those of $A_k{}^*$ and $B_k{}^*$ found by the analysis the values of $P_k{}'$, $R_k{}'$, $P_k{}''$ and $R_k{}''$ are computed and are given together with the corresponding values of $A_k{}^*$ and $B_k{}^*$ in table 2.

Table 2

Ordinates	Coefficients of creep curve ' remainder '		Coefficients of J'		Coefficients of J''	
	Cosine	Sine				
k	$A_k{}^*$	$B_k{}^*$	$P_k{}'$	$R_k{}'$	$P_k{}''$	$R_k{}''$
1	0·9528	−3·6405	−1472·0	3562·2	1308·9	540·89
2	0·5501	−0·2060	− 636·16	74·973	48·567	412·10
3	0·1157	0·0683	− 107·58	−117·935	−96·688	88·197
4	−0·0390	0·0903	83·397	− 96·874	−88·386	76·090
5	−0·0524	0·0068	74·232	9·784	9·381	71·172
6	−0·2390	0·0006	35·029	5·479	5·373	34·352
7	−0·0172	0·0068	27·767	12·691	12·576	27·517
8	0·0026	0·0113	− 6·589	+ 19·348	−19·265	6·561
9	0·0051	0·0040	− 11·124	− 4·532	− 4·523	11·102

As only the 'remainder' of the creep is expressed by the Fourier series, expressions (1) and (2) give only the periodic parts of J' and J''. The complete functions consist of the cosine and sine terms obtained by synthesizing $P_k{}'$ and $R_k{}'$, again using the Danielson and Lanczos method, plus the linear term removed and a constant term.

The linear term for J' is

$$P' = \frac{(\tau_N + 12{\cdot}8)}{25{\cdot}6}\{\xi_{+32} - \xi_{-32}\}$$

where τ_N is the value of $\ln t$ at the ordinate N, ξ_{+32} and ξ_{-32} are the creep ordinates at the long and short time ends of the domain respectively.

The constant was chosen so that when all the J' terms were added at the extreme long time end of the domain they equalled the corresponding creep ordinate.

In the same manner, a constant term $P'' = \frac{1}{2}\pi(\xi_{+32} - \xi_{-32})/25{\cdot}6$ was added to the cyclic components of J''.

From the values of $J'(\omega)$ and $J''(\omega)$ one can derive all the other dynamic quantities; thus $G' = J'/(J'^2 + J''^2)$, $G'' = J''/(J'^2 + J''^2)$, $\tan\delta = J''/J'$, where G', G'' are the real and imaginary parts of the dynamic shear modulus and δ is the phase angle between the stress and strain.

In the Alfrey approximation, the retardation spectrum is given by the derivative, with respect to ln (time), of the creep curve. On the other hand, the same type of approximation makes the retardation spectrum the derivative of the real component of the dynamic compliance. Thus, applying the Alfrey approximation twice, one has the result that the creep curve is identical with J'. Again with the same type of approximation, the imaginary component of the dynamic compliance and the retardation spectrum are related by a constant factor. Thus in practice approximate values of J' and J'' are obtained very simply from a creep function; $J'(\ln\omega)$ is taken to be identical with $\xi(\ln t)$, and $J''(\ln\omega)$ is identified with $\frac{1}{2}\pi\, d[\xi(\ln t)]/d(\ln t)$. If the retardation spectrum varies rapidly with changes of $\ln t$ the approximations are poor but if, as with polythene, the spectrum is flat over many decades the approximation can be expected to work well.

§ 4. Discussion of Results

In figure 4 the values of G' determined from the creep curve are shown as a function of $\log_{10}$ (time). The same figure also shows both the approximate values of G' as determined by Alfrey's method and some values determined directly from dynamic measurements made on the same specimen. Figure 5 shows a similar set of curves for $\tan\delta$. Inspection of both figures shows that even with polythene, which has a very broad retardation spectrum, there is a significant difference between the approximate results and those derived by Fourier analysis. The general effect of the approximation is to move the dynamic quantities along the time axis. Since $\tan\delta$ equals the ratio of J'' to J' the errors in this curve are less than in the case of the moduli.

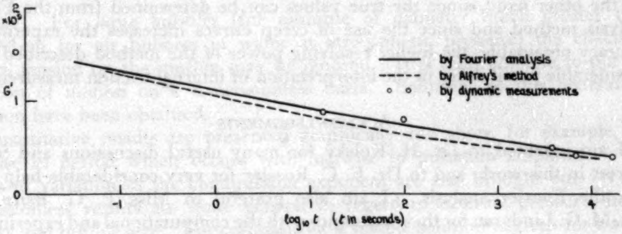


Figure 4. Shear modulus–$\log_{10}$ (time).

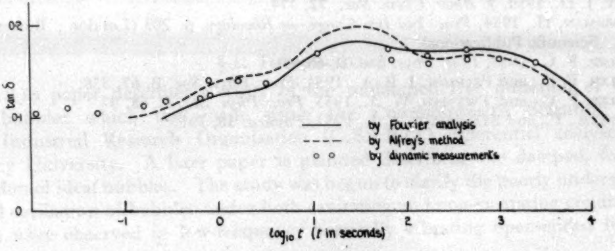


Figure 5. Tangent of loss angle – $\log_{10}$ (time, secs.).

Further, since the ($\tan\delta$, time) curve of polythene is very flat the effect of the shift along the time axis is very small; in the presence of a sharp peak the effects would be much more pronounced. Nevertheless, the simple approximation does give values of $\tan\delta$ which differ from the true values by amounts of the same magnitude as the experimental errors of the equivalent dynamic measurements. The points due to direct dynamic measurements and the curve for $\tan\delta$ derived from creep results deviate by up to 15%. The larger part of this deviation is due to the inherent inaccuracies of the oscillatory experiments.

§ 5. CONCLUSIONS

This work shows that, by using a Fourier analysis technique, creep curves can be used to provide accurate values of the dynamic visco-elastic properties of polymers under conditions where direct dynamic measurements are impracticable. The accuracy of the derived dynamic quantities is limited only by the experimental errors to which the creep measurements are subject. For substances with flat spectra the work furthermore suggests that even when using approximate methods of transformation the accuracy of the $\tan\delta$ values is equivalent to that obtainable from direct dynamic experiments. In the present example the shear modulus and internal friction found by Alfrey's method differ from the true value but since polythene has a very flat retardation spectrum the errors are small and might be unimportant in many practical applications. On the other hand, since the true values can be determined from the Fourier analysis method and since the use of creep curves increases the experimental accuracy obtainable, the higher resolving power of the method described offers considerable possibilities in the interpretation of internal friction measurements.

ACKNOWLEDGMENTS

I am indebted to Dr. H. Kolsky for many useful discussions and for his interest in this work, and to Dr. F. C. Roesler for very considerable help given with the Fourier analysis. I am also grateful to Miss P. G. Bailey and Mrs. M. G. Lindgren for their assistance with the computational and experimental parts of the work.

REFERENCES

ALFREY, T., 1948, *Mechanical Properties of High Polymers* (New York : Interscience).
DANIELSON, G. C., and LANCZOS, C., 1942, *J. Franklin Inst.*, **233**, 365.
FERRY, J. D., 1950, *J. Amer. Chem. Soc.*, **72**, 374.
LEADERMAN, H., 1954, *Proc. 2nd Int. Congr. on Rheology*, p. 203 (London : Butterworths Scientific Publications).
ROESLER, F. C., 1955, *Proc. Phys. Soc.* B, **68**, 89.
ROESLER, F. C., and PEARSON, J. R. A., 1954, *Proc. Phys. Soc.* B, **67**, 338.
ROESLER, F. C., and TWYMAN, W. A., 1955, *Proc. Phys. Soc.* B, **68**, 97
SCHWARZL, F., and STAVERMAN, A. J., 1952, *Physica*, **18**, 791.

Undamped Free Pulsations of an Ideal Bubble

By R. B. ROBINSON† and R. H. BUCHANAN‡

† Department of Electrical Engineering, University of Sydney

‡ Department of Chemical Engineering, New South Wales University of Technology

MS. received 20*th December* 1955, *and in revised form* 6*th March* 1956

Abstract. For large bubbles (for example of diameter much greater than 3×10^{-4} cm for air bubbles in water at atmospheric pressure) the effects of surface tension on the motion may be ignored. It is then possible to put the equation of motion on a dimensionless basis. Solutions of this generalized equation have been obtained.

Quantitative results are presented graphically and show, for example, the increase in period which occurs with increase in pulsation amplitude. The effects of variation of the compression exponent are also brought out. These dimensionless results are quite general in regard to external pressure, liquid density and bubble size. For comparison, a sample solution is given for a small bubble with allowance for surface tension.

Discussion covers both the correlation of the results with available experimental evidence, and also various implications of the results, for example, in regard to forced pulsation.

§ 1. INTRODUCTION

THIS paper describes a study of the undamped free pulsation of ideal bubbles which was made using the Commonwealth Scientific and Industrial Research Organisation (C.S.I.R.O.) differential analyser at Sydney University. A later paper is planned describing the damped, forced pulsation of ideal bubbles. The study was begun to clarify the poorly understood forced oscillation of bubbles under both cavitating and non-cavitating conditions which were observed in low-frequency vertically vibrating open-ended liquid columns.

Various investigators have considered the small-amplitude, undamped, free pulsation of ideal bubbles. For example, Minnaert (1933) measured the frequency of sound produced by bubbles of varying sizes in rising through quiescent liquids. He found that this frequency corresponded closely with the theoretical adiabatic pulsating frequency for very small amplitudes:

$$\omega = \frac{1}{R_0}\left\{\frac{3\gamma P_A}{\rho}\right\}^{1/2} \qquad \ldots\ldots(1)$$

where ω is the frequency in rad sec^{-1}, R_0 the equilibrium bubble radius in cm, γ the ratio of specific heats of gas at constant temperature and constant volume, P_A the external pressure in dyn cm^{-2} and ρ the liquid density in g cm^{-3}.

This basic equation was modified as follows by Smith (1935) to consider the effect of surface tension S dyn cm^{-1}:

$$\omega = \frac{1}{R_0}\left\{\frac{3\gamma P_A + 2S/R_0}{\rho}\right\}^{1/2}. \qquad \ldots\ldots(2)$$

As pointed out by J. M. Richardson (see Briggs, Johnson and Mason 1947), this equation is incorrect and should be written as follows (see later theoretical treatment):

$$\omega = \frac{1}{R_0}\left\{\frac{3\gamma(P_A + 2S/R_0) - 2S/R_0}{\rho}\right\}^{1/2}. \qquad \ldots\ldots(3)$$

All the above equations assume sinusoidal pulsation, which, due to spherical divergence and convergence, is only approached with very small amplitudes. Little work has been done, however, on large-amplitude pulsation due to the inherent mathematical difficulties.

The object of this paper is to elucidate theoretically this effect of large-amplitude pulsations on such things as the natural frequency and the percentage of the cycle expanded. Surface tension and heat transfer to the surrounding liquid are also considered. The work of Noltingk and Neppiras (1950, 1951) suggested the approach to the problem.

§ 2. EQUATION OF MOTION

The equation is written for a bubble of radius R cm in an infinite, incompressible, inviscid liquid. The following effects are considered.

(i) Surface tension of the liquid S dyn cm^{-1} causes an inward pressure $2S/R$ dyn cm^{-2}.

(ii) The gas content of the bubble causes an outward pressure P, which is given by the equation $PR^{3n} = P_0 R_0^{3n}$ where P_0 and R_0 are the equilibrium values and n is the compression exponent. This exponent depends on the nature of the gas and of the heat transfer, which is assumed to be reversible in this undamped case.

(iii) The external pressure at infinity P_A acts inwards and is here assumed constant. The pressure in the liquid at radius R is denoted by P_R.

(iv) The moving liquid mass is easily shown to possess the total kinetic energy E ergs at any time:

$$E = 2\pi\rho R^3 (dR/dt)^2. \qquad \ldots\ldots(4)$$

The equation of motion is then obtained from the energy relation

$$(P_R - P_A)(4\pi R^2\, dR) = dE \qquad \ldots\ldots(5)$$

and is

$$2\left\{\left(P_A + \frac{2S}{R_0}\right)\left(\frac{R_0}{R}\right)^{3n} - P_A - \frac{2S}{R}\right\} = 3\rho\left(\frac{dR}{dt}\right)^2 + 2\rho R\frac{d^2R}{dt^2}. \qquad \ldots\ldots(6)$$

This equation neglects the kinetic energy of the gas, and vaporization, while the effect of the static pressure head can be approximated by an adjustment of P_A.

It is interesting to consider the linearized equation for small pulsations. With $R = R_0 + \epsilon$ and $\epsilon/R_0 \ll 1$, it is possible to simplify the equation of motion to the following linear form by dropping second order terms

$$\frac{d^2\epsilon}{dt^2}+\epsilon\left\{3n\left(P_{\mathrm{A}}+\frac{2S}{R_0}\right)-\frac{2S}{R_0}\right\}\left(\frac{1}{\rho R_0^{2}}\right)=0. \qquad \ldots\ldots(7)$$

The solution of this equation is sinusoidal with a frequency agreeing with Richardson's equation, (3), rather than Smith's, (2). By linearizing the radius–volume relationship for small variations, it follows that both the radius and volume vary sinusoidally in the limiting case.

Referring again to the full equation (6), the effect of surface tension prevents this equation being written in a more general form. However, this effect is progressively less important with increase in bubble radius R and can be ignored when $R \gg 2S/P_{\mathrm{A}}$. When this obtains it is possible to generalize the equation of motion by using the dimensionless variables $\beta = R/R_0$ and $\tau = t/T_0$ where T_0 is the limiting period for small pulsations, $2\pi R_0(\rho/3nP_{\mathrm{A}})^{1/2}$ seconds. The equation of motion then becomes

$$\frac{8\pi^2}{3n}\{\beta^{-3n}-1\}=3\left(\frac{d\beta}{d\tau}\right)^2+2\beta\frac{d^2\beta}{d\tau^2}. \qquad \ldots\ldots(8)$$

In the case of air bubbles in water at atmospheric pressure, this holds when $R \gg 1{\cdot}5\times10^{-4}$ cm as is generally the case with visible bubbles.

§ 3. SOLUTIONS

It will be noticed in equation (8) above that the only variables are β, n and τ, since P_{A}, ρ and R_0 disappear in deriving the dimensionless form. Consequently, to cover all feasible cases this equation was solved only for a range of values of n treated as a parameter and for various initial conditions of β and $d\beta/d\tau$ with each value of n.

To facilitate the computing, integral values of $3n$ were taken, namely 3, 4 and 5. These values cover the range generally encountered, as is discussed later.

The initial value β_0 of β was always taken at one and a range of values of initial slope $(d\beta/d\tau)_0$ was used. This gave full generality.

Solutions were obtained using the C.S.I.R.O. mechanical differential analyser. The general procedure for these free pulsations was to set initial conditions and then make the machine plot β against τ.

In high amplitude pulsations, the large ranges of the variables encountered caused difficulties in retaining accuracy on the differential analyser. It was often necessary to change scaling factors during a run for this reason. For the same reasons, and as normal procedure, the solutions were checked as closely as possible by independent means. For example, at points of contraflexure of the (β, τ) curve $d^2\beta/d\tau^2$ is zero, making it possible to check the computer closely for compliance with the equation. Another important check was given by the calculated maximum and minimum values of β for a given initial $d\beta/d\tau$. The relationship, in general terms, between the initial conditions β_0 and $(d\beta/d\tau)_0$ and the extremes of β, β_m, was derived from energy considerations and is

$$\frac{3n}{8\pi^2}\left(\frac{d\beta}{d\tau}\right)_0^2=\int_{\beta_0}^{\beta_m}(1-\beta^{-3n})\beta^2\,d\beta. \qquad \ldots\ldots(9)$$

The integral is easily evaluated for any value of n, giving an equation in β_m solved by trial and error.

For comparison with the above solutions, the full equation of motion (6) including surface tension, was solved for a sample case as follows: $R_0 = 6 \times 10^{-4}$ cm, $S = 75$ dyn cm^{-1}, $\rho = 1$ g cm^{-3}, $P_A = 10^6$ dyn cm^{-2} and $n = 1$ using various initial conditions of dR/dt. It will be clear from equation (6) that this set of solutions does not have the same generality as the previous ones.

§ 4. DISCUSSION

The results are presented in various ways in figures 1 to 6. Figure 1 shows maximum and minimum values of β, $\beta_{\max}$ and $\beta_{\min}$, plotted against $(d\beta/d\tau)_0$ when n is 1, 4/3 and 5/3. These were calculated as described in the previous section and are in good agreement with the machine values. It is clear that the higher the value of n, the less is the pulsation ratio Z $(= \beta_{\max}/\beta_{\min})$ for a given value of $(d\beta/d\tau)_0$. In other words, the pulsation ratio of any bubble for a given amount of pulsation energy decreases as n rises.

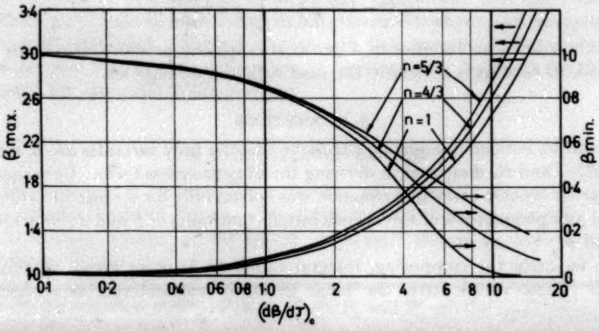


Figure 1. Variation of $\beta_{\max}$ and $\beta_{\min}$ with $(d\beta/d\tau)_0$ for $n = 1$, 4/3, 5/3.

The values of $\beta_{\max}$ and $\beta_{\min}$ for the same three values of n are plotted against each other in figure 2. These curves show that the smaller the value of n, the less is $\beta_{\min}$ for a given $\beta_{\max}$. Also the effect of surface tension is shown to be similar to increasing n.

A typical set of machine curves of β plotted against τ is given in figure 3 for an n of 4/3. It is noted that with rising amplitude there is a steady increase in period, in the deviation from simple harmonic motion and in the percentage of the cycle in the expanded phase. Further, it may be seen that with rising amplitude the contraflexure point of maximum radial velocity occurs ever nearer the minimum.

Figure 4 shows sets of typical derived radius–time curves for a particular bubble using values of n of 1, 4/3 and 5/3, with $R_0 = 0{\cdot}5$ cm, $P_A = 10^6$ dyn cm^{-2}, $\rho = 1$ g cm^{-3} and $(dR/dt)_0 = 300$ and 2500 cm sec^{-1}. The necessary data for constructing these curves were obtained from the previous equations and the various figures. This plot shows the decrease in $R_{\max}$ and increase in $R_{\min}$ at constant pulsating energy with rising n which was mentioned earlier in connection with figure 1. In addition, it may be seen that n has most effect on the period when the pulsation amplitudes are small.

The effect of the pulsation amplitude on the period is illustrated by figure 5 in which τ_p, the period in terms of τ, is plotted against Z. It may be noticed that small pulsation ratios have little effect on the period and that the greater the value of n, the greater is the effect at the higher ratios. The effect of surface

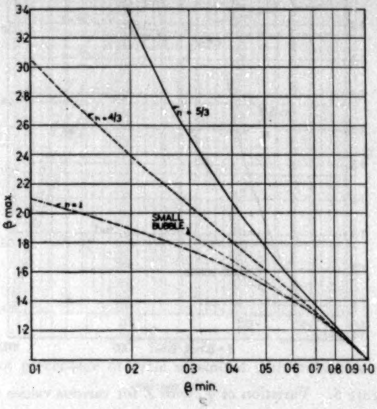
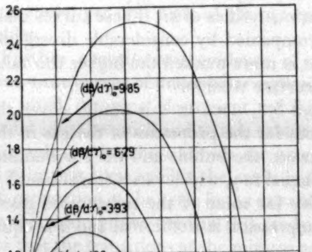


Figure 2. Variation of $\beta_{\max}$ with $\beta_{\min}$ for various values of n.



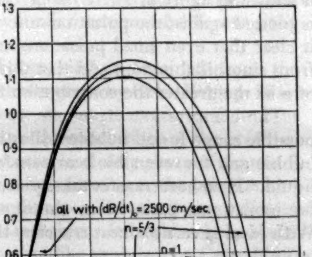


Figure 3. Variation of β with τ for various values of $(d\beta/d\tau)_0$ when $n=4/3$. C is a point of contraflexure.

Figure 4. Variation of R with t for different values of n and $(dR/dt)_0$.

tension on the period of the $6{\cdot}0\times10^{-4}$ cm bubble is once again seen to be similar to that produced by increasing n.

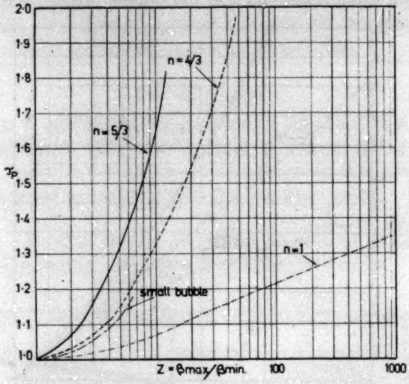


Figure 5. Variation of τ_p with Z for various values of n.

The deviation from simple harmonic motion with rising pulsation amplitude is shown in figure 6, where the percentage of the cycle in the expanded condition is plotted against the pulsation ratio for various values of n. These curves make it clear that even small pulsations are accompanied by considerable divergence from sinusoidal motion and that this effect is more marked the higher the value of n or the greater the contribution from surface tension.

The compression exponents, 1, 4/3 and 5/3, used in this study, cover the possible range for all bubbles, allowing both for the properties of the gas in the bubble and for reversible heat transfer between the bubble and the surrounding liquid. If no heat transfer takes place, n is equal to γ, which varies from about 5/3 for monatomic gases down almost to unity for some of the polyatomic gases. With ideal reversible heat transfer, the compression is isothermal and n is equal to unity.

With real bubbles, heat transfer will occur and will be partially irreversible, thereby damping the motion. This heat transfer causes the pressure–volume relationship to differ from the ideal cases above. It usually remains close to the ideal form where PV^n is constant but with n intermediate between γ and 1. The experimental evidence in the case of small pulsations is that n departs only slightly from γ and the compression is almost adiabatic (Minnaert 1933). For somewhat larger pulsations, the experiments of Meyer and Tamm (1939) again show n to approximate γ. However, the accuracy of this work was necessarily lower and a small deviation would not be apparent. For very large pulsations, while no experimental evidence is available, it is to be expected that n departs further from γ, and that damping of the motion is greater, owing to the emergence of energy losses which increase more with either temperature or velocity than would be the case in a linear system. Examples of such losses are those due to turbulence, heat radiation and sound radiation. The losses due to these effects may become large under cavitating conditions.

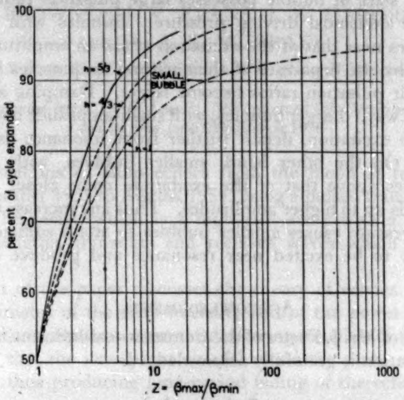


Figure 6. Variation of percentage of cycle expanded with pulsation ratio *Z* for various values of *n*,

The value of *n* is thus close to γ and is relatively independent of pulsation ratio. It appears that the value of *n* is also relatively independent of bubble size. This is so because of the theoretical constancy of the temperature-proportional heat transfer per unit mass of gas per cycle for similar motions of different sized bubbles.

Further experimental evidence of a different kind is obtained from the radius–time curves of underwater explosion bubbles given by Cole (1948, p. 271). The small damping decrement exhibited here shows that irreversible heat transfer must be small and to this extent confirms the conclusions of the previous paragraphs. Also these curves are noted to resemble closely those of figures 3 and 4 in that they exhibit the same high radial velocities near the minima and high proportion of the cycle in the expanded conditions. This similarity indicates that the assumptions used in this study are reasonable. It has also been made clear that the transitions at the minima are not cusp-like as shown by Cole (1948) but rather that they have an appreciable radius of curvature which steadily decreases with rising *Z*.

Since the ratio of the pressure *P* within the bubble to the equilibrium pressure P_A is equal to $(1/\beta)^{3n}$, it follows that extremely high pressures are attained near the minima with large pulsation ratios. These pressures would of course be modified considerably in reality by mass-transfer across the liquid–gas interface. With large pulsation amplitudes, considerable transfer of dissolved gas as well as liquid vaporization and condensation occur in each cycle. The experimental observation that there is always a net transfer into a bubble undergoing forced pulsation indicates that the long time in the expanded phase with large surface area and small driving forces is more effective than the short time in the contracted phase with small surface area and large driving forces.

As would be expected, the increase in period with rising pulsation ratios is an important factor in the forced pulsation of bubbles. Because of this factor, a broad range of sizes of bubble possesses large pulsation ratios in any forced motion due to a sinusoidal driving pressure. Bubbles with small-amplitude natural frequencies near that of the excitation attain an amplitude magnification which is limited by the departure of their natural frequencies from that of the excitation, as their pulsation ratios become larger. Damping also is a limiting factor in the usual way. Larger bubbles, with small-amplitude natural frequencies below that of the excitation, depart further from resonance as their pulsation ratios increase. On the other hand, smaller bubbles, with small-amplitude natural frequencies above that of the excitation, come closer to resonance as their motion builds up to higher amplitudes. Thus any increase in the amplitude of the exciting pressure causes smaller bubbles to attain sufficient reduction in natural frequency to be excited near resonance and produce large amplitude magnifications.

ACKNOWLEDGMENTS

The assistance of Mr. J. Foster of the Commonwealth Scientific and Industrial Research Organisation is gratefully acknowledged.

REFERENCES

BRIGGS, H. B., JOHNSON, J. B., and MASON, W. P., 1947, *J. Acoust. Soc. Amer.*, **19**, 664.
COLE, R. H., 1948, *Underwater Explosions* (Princeton: University Press).
MEYER, E., and TAMM, K., 1939, *Akust. Z.*, **4**, 145.
MINNAERT, M., 1933, *Phil. Mag.*, **16**, 235.
NEPPIRAS, E. A., and NOLTINGK, B. E., 1951, *Proc. Phys. Soc.* B, **64**, 1032.
NOLTINGK, B. E., and NEPPIRAS, E. A., 1950, *Proc. Phys. Soc.* B, **63**, 674.
SMITH, F. D., 1935, *Phil. Mag.*, **19**, 1147.

Radio Echoes from the Moon

By I. C. BROWNE,[†] J. V. EVANS, J. K. HARGREAVES [‡]
and W. A. S. MURRAY [§]
Jodrell Bank Experimental Station, University of Manchester

Communicated by A. C. B. Lovell; MS. received 6th December 1955,
and in final form 6th March 1956

Abstract. Observations of radio echoes from the moon at transit have been made from September 1953 to August 1955, using a pulsed transmitter operating at 120 Mc/s. The aerial, which is a high-gain array with a physical area of 40 square wavelengths, transmits and receives waves which are horizontally polarized.

The first part of the paper discusses the theory of echoes from the moon, and includes estimates of the echo intensity and of the power spectrum of the echoes for several assumptions about the scattering function of the lunar surface. It is also shown that the ionosphere should rotate the plane of polarization of the radio waves, thus producing long-period fading of the echoes, and that the angle of rotation is a measure of the total electron content of the ionosphere.

The second part of the paper presents the results of the observations. The observed echo intensity, in conjunction with the observed power spectra of the echoes, is consistent with the assumption that the moon reflects radio waves in such a way that its projected disc appears uniformly 'bright', as is the case for light waves. The power reflection coefficient must be somewhat less than 0·1.

The observed echoes are subject to rapid fading. This is found to be due to the moon's libration. The libration varies with the position of the moon in its orbit, causing a corresponding change in the rate of fading. Experimental proof is obtained that the slow fading of moon echoes previously reported by other workers is due to the rotation of the plane of polarization of the waves by the ionosphere. By observing the slow fading on two frequencies with a separation of 1·5 Mc/s, it is possible to measure the total angle by which the plane of polarization is rotated and hence the total electron content of the ionosphere. Some preliminary results appear to indicate that the parabolic half-thickness of the F2 region may be a more reliable measure of the extent of the ionosphere than has sometimes been supposed.

§ 1. Introduction

Radio echoes from the moon have been obtained by several workers. Bay (1946) working on a frequency of 120 Mc/s first claimed to have detected echoes, but he used a chemical integrator to confirm their presence, and was unable to make any measurements of the intensity, or other characteristics of the echoes. The first results using more conventional equipment were obtained in America by De Witt and Stodola (1949), who observed at 111·5 Mc/s at moonrise and moonset. They reported a maximum signal strength 7 dB less than expected although, as will be seen, the maximum observed signal strength does not give a reliable indication of the mean echo intensity. Unexpected fluctuations in signal strength were observed to have a period of several minutes and were attributed to anomalous ionospheric refraction.

[†] Now at The Physics Department, University of Sheffield.
[‡] Now at The Radio Research Station, Slough, Bucks.
[§] Now at The Radar Research Establishment, Malvern, Worcs.

A similar unexplained fading was observed by Kerr, Shain and Higgins (1949) who used frequencies close to 20 Mc/s and again observed at moonrise and moonset. They were able to show that a rapid component of the fading, having a period of a few seconds, could be explained by assuming that the moon reflects as though it were an assembly of random scattering points in relative motion caused by the moon's libration. Like De Witt and Stodola, they attributed the slow component of the fading to ionospheric refraction. The greatest mean echo intensities observed were about 10 dB lower than expected.

Sulzer, Montgomery and Gerks (1952) observed at a frequency of 418 Mc/s, and measured maximum signal strengths about 7 dB less than expected on the basis of a Lommel–Seeliger scattering law (see § 2.1).

The present paper describes work carried out at a frequency of 120 Mc/s with the moon in transit. The mean echo intensity is compared with theoretical estimates, and the nature of the rapid fading of the echoes is investigated in relation to various laws of scattering from the moon's surface. The paper also describes the investigation of the slow fading of the moon echoes with respect to the rotation of the plane of polarization of the radio waves during their passage through the ionosphere. The effect is used to estimate the total electron content of unit cross section between the observer and the moon.

§ 2. THEORETICAL CONSIDERATIONS

2.1. *The Intensity of the Echoes*

The echo power P_r received from a spherical target of radius a is given by the well-known radar equation:

$$P_r = P_t G A \frac{a^2}{16\pi R^4} g\rho$$

where P_t is the peak transmitted power, G the power gain of transmitting aerial over an isotropic radiator, A the effective collecting area of the receiving aerial, R the range of the target, g the directivity for the target (see below) and ρ the power reflection coefficient for the material of which the target is composed. The weakest echo which can be detected is limited by the presence of noise radiated from outside sources and arising in the receiver itself. When the external noise is weak compared with the receiver noise it is convenient to measure the echo intensity in terms of the receiver noise power P_n, and we have:

$$\frac{P_r}{P_n} = \frac{P_t}{n_r \boldsymbol{k} T_0 b} G A \frac{a^2}{16\pi R^4} g\rho$$

where n_r is the receiver noise factor, $\boldsymbol{k}$ Boltzmann's constant ($=1{\cdot}38 \times 10^{-23}$ w deg^{-1}), $T_0 = 290$°K and b is the energy bandwidth of the receiver, defined as

$$\frac{1}{V_0^2}\int_0^\infty V^2(f)\,df$$

where $V(f)$ is the overall voltage gain at frequency f. In this form of the equation, it is assumed that all the transmitter power is radiated at frequencies within the receiver bandwidth.

It can be seen that the signal-to-noise ratio P_r/P_n can be increased by narrowing the receiver bandwidth, although in practice there is no further improvement if the bandwidth is narrower than the reciprocal of the transmitter pulse-length.† The accuracy of measurement of the range R of the target is given approximately by $\delta R = \mathbf{c}/2b$ where $\mathbf{c}$ is the velocity of light. Since the distance of the moon is measurable optically to one part in 10^6, the loss of accuracy in the measurement of range caused by using a narrow bandwidth is of no consequence.

If the moon were a smooth sphere very large compared with the wavelength, its equivalent echoing area, defined as the geometric cross section of an isotropic scatterer giving the same echo intensity as the actual target, would be equal to $\rho\pi a^2$ (Norton and Omberg 1947). Spheres which are not smooth will have a different equivalent echoing area which can be written $g\rho\pi a^2$, where g is the directivity of the target.

Two special cases have been studied in some detail. If the surface irregularities are about one wavelength in size, Lambert's law will apply, and the directivity g can be shown to equal 8/3 (Grieg, Metzger and Waer 1948). In this case, the moon will not reflect radio waves uniformly from each element of its projected disc, but will show limb 'darkening' according to a cosine law.

When the average irregularities are large compared with the wavelength and are within a few orders of magnitude of the radius of the target, as in the case of a mountainous surface, each hill near the limb will contribute favourably in the line-of-sight, and at the same time the far side of the hill and any flat region beyond, which would re-radiate mainly away from the line-of-sight, will be screened from view. In an ideal case of this kind, the echo power from every element of the projected disc is the same, so that the disc appears uniformly 'bright' (Lommel–Seeliger scattering). The moon approximates to this form of scattering for light and Kerr and Shain (1951) have deduced an optical directivity of 5·7 from figures for the variation with phase of the optical brightness of the moon (Russell, Dugan and Stewart 1945, p. 173). In the absence of other evidence, we shall assume that the value of 5·7 gives the directivity of the moon for radio waves of wavelength 2·5 m.

In order to estimate the power reflectivity ρ of the moon, it is necessary to know the chemical constitution of its surface layers to a depth of the order of a wavelength. Most evidence (Russell, Dugan and Stewart 1945, p. 173, Piddington and Minnett 1949) suggests that the moon's surface consists of dry rock with a covering of dust as least a few millimetres thick. Measurements of the power reflectivity of dry terrestrial rocks, sand and dust lead to values of ρ of the order of 0·1 at frequencies greater than 10 Mc/s, and it seems likely that for the extremely dry conditions which must prevail on the moon a value of $\rho=0{\cdot}1$ will set an upper limit. It should be remembered that the composition of the moon's surface probably varies from place to place.

---

† In the moon echo work the optimum pulse length is also governed by other conditions If the moon reflects as a rough surface then the returned signal will be proportional to the amount of surface area illuminated. The radar depth of the moon is 11·6 msec, and any pulse shorter than this will be returned as one at least 11·6 msec long. Thus the transmitted pulse must be at least equal to this value if the whole of the lunar surface is to be effective in scattering. On the other hand if the moon behaves as a smooth sphere then the maximum returned signal will be achieved as soon as the first few Fresnel zones of the lunar surface are illuminated.

Adopting the values $g=5{\cdot}7$, $\rho=0{\cdot}1$, $a=1{\cdot}74\times10^8$ cm, the moon's expected equivalent echoing area is $g\rho\pi a^2=5{\cdot}41\times10^{16}\,\text{cm}^2$. The observations (§ 4.1) show that the true value *may* be less by a factor of not more than three times.

2.2. *The Doppler Shift and the Fast Fading of Echoes*

Because the moon's orbit is an ellipse, with a mean eccentricity of 0·055, the moon has, in general, a component of velocity towards the centre of the earth. In addition, an observer on the surface of the earth will have a velocity component in the line-of-sight, due to the rotation of the earth. The magnitude of this velocity will vary with the latitude of the observer and with the hour-angle of the moon. The resultant of these two velocities produces a Doppler shift in the frequency of echoes from the moon, and the frequency to which the radio receiver is tuned must be adjusted accordingly. At a frequency of 120 Mc/s, the Doppler shifts are of the order of ± 50 c/s for observations when the hour-angle of the moon is less than 30 minutes of time.

There is another important consequence of the motions of the moon. If the moon is regarded as an assembly of scattering elements distributed randomly over its surface, the intensity of the echo at any instant is determined by the resultant of the signals scattered from each element. The intensity will be constant only if the phases of the individual signals do not vary relative to each other; that is, if the moon always presents the same face to the observer. In fact, as is well known, the moon librates, or rocks about an axis which itself changes with time. The libration produces changes in the phases of the signals from the different scatterers, so that the echo fades at a rate proportional to the rate of rocking. The probability that the signal will have a given amplitude is given by the Rayleigh law, and is independent of the total number of scatterers, provided that this is large (Lawson and Uhlenbeck 1950, p. 47).

The libration has four main components. These are described in the Appendix, where it is shown that, for the present purpose, only three are important. These are the librations in latitude l_β, in longitude l_L and the diurnal libration l_D. Both l_β and l_L are tabulated in the *Nautical Almanac* in the form of the total angle of libration for each day of the year. The diurnal libration may be calculated with sufficient accuracy from the formula in the Appendix. The total libration l_T is obtained by addition of l_L, l_β and l_D as described in the Appendix. The calculation of the power spectrum of the echoes for a given value of l_T is outlined in the next section.

2.3. *The Power Spectrum of Moon Echoes*

In figure 1 (*a*) the moon is represented by a sphere of radius a which appears to be rotating with an angular velocity l_T about an axis OO′ perpendicular to the line-of-sight. Coordinates are measured from the point where the line-of-sight meets the sphere. A point on the sphere with latitude β and hour-angle ϕ, will have a component of velocity towards the observer given by

$$v=al_T\cos\beta\cos\alpha\sin\phi.$$

This shows that the sphere can be replaced by its projected disc (figure 1 (*b*)) and the position of the point measured in cartesian coordinates xy, where $x = a \sin\phi \cos\beta$, and $y = a \sin\beta$. The line-of-sight velocity is now given by

$$v = xl_{\mathrm{T}} \cos\alpha \simeq xl_{\mathrm{T}}$$

since for the moon α is always less than 15′. Thus the moon can be divided into strips of constant x, each of which returns signals differing from the frequency of the transmitted wave by an amount $f = 2xl_{\mathrm{T}}\lambda$, where λ is the radio wavelength.

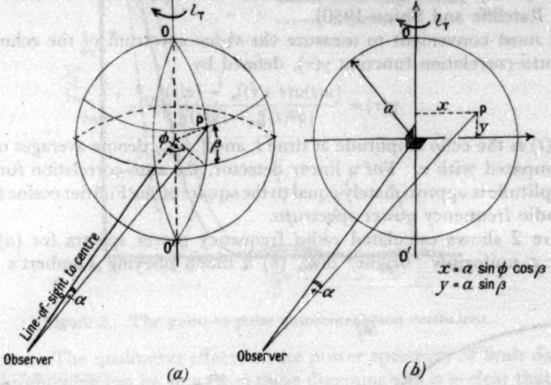


Figure 1. (*a*) The sphere representing the moon on which a point P has hour angle ϕ and latitude β. (*b*) The equivalent disc, where the point P may be represented by its Cartesian coordinates.

If f_0 is the Doppler shift produced by either limb of the moon, then since in the present equipment $\lambda = 2{\cdot}5$ metres

$$f_0 = l_{\mathrm{T}} \times 1{\cdot}4 \times 10^6 \text{ (c/s)}.$$

When the total libration l_{T} is at its maximum, the resulting Doppler broadening of the echo power spectrum is only of the order of ± 2 c/s. Although this is not detectable directly as a broadening of the echo power spectrum, it results in a fluctuation of the echo amplitude from pulse to pulse. Thus by comparing the relative amplitudes of successive pulses, the radio frequency power spectrum and thence the lunar scattering law may be deduced.

If the moon is regarded as an assembly of randomly placed scatterers, the power returned from any strip of constant x is equal to $s(x, y)\,dx\,dy$, where $s(x, y)$ is the appropriate scattering function relating the back-scattered signal to the angle of incidence. The power spectrum of the returned echo is thus given by

$$P(f)\,df = \int s(f\lambda/2l_{\mathrm{T}}, y)\,\frac{\lambda}{2l_{\mathrm{T}}}\,df\,dy$$

where $P(f)\,df$ measures the power returned in the interval df at a frequency differing from the frequency of the transmitter by f. The upper and lower limits of y are, of course, related to f by the equation $y^2 + (f\lambda/2l_{\mathrm{T}})^2 = a^2$.

The integral can be evaluated for a given scattering law $s(x, y)$. The simplest forms of scattering law are those which, like Lambert's law, show symmetry about the centre of the moon's projected disc.

$P(f)df$ gives the radio frequency power spectrum of the returned echo, and this is not generally the same as the video frequency power spectrum, because cross-modulation terms are introduced at the detector. Lawson and Uhlenbeck (1950, p. 59) show how to calculate the video frequency spectrum from the radio frequency spectrum for both linear and square-law detectors (see also Booker, Ratcliffe and Shinn 1950).

It is most convenient to measure the video spectrum of the echo in terms of the auto-correlation function $\gamma(\tau)$, defined by

$$\gamma(\tau)=\frac{[z(t)z(t+\tau)]_{a}-[z(t)]_{av}^{2}}{[z^{2}(t)]_{av}-[z(t)]_{av}^{2}}$$

where $z(t)$ is the echo amplitude at time t, and $[\ldots]_{av}$ denote averages over times large compared with τ. For a linear detector, the auto-correlation function for echo amplitude is approximately equal to the square of the Fourier cosine transform of the radio frequency power spectrum.

Figure 2 shows calculated radio frequency power spectra for (*a*) a moon showing a uniformly 'bright' disc, (*b*) a moon obeying Lambert's law, and

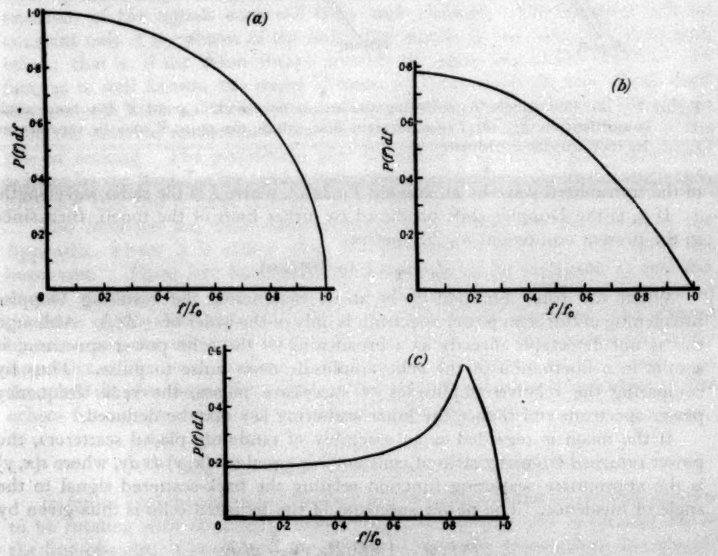


Figure 2. The radio frequency power spectrum for (*a*) Lommel–Seeliger scattering, (*b*) Lambert scattering, (*c*) a limb-brightened disc, as described in the text, where f is the Doppler shift produced by a given point on the disc and f_0 that produced at the limb.

(*c*) a moon whose projected disc appears as a uniformly bright annulus, with outer radius equal to that of the projected disc and area one-third that of the disc. Figure 3 shows the video auto-correlation functions corresponding to these spectra. The qualitative effects of the power spectrum of limb darkening and limb brightening can be seen from these diagrams and it is clear that a study of the rapid fading of moon echoes can help in evaluating the law of scattering for radio radiation.

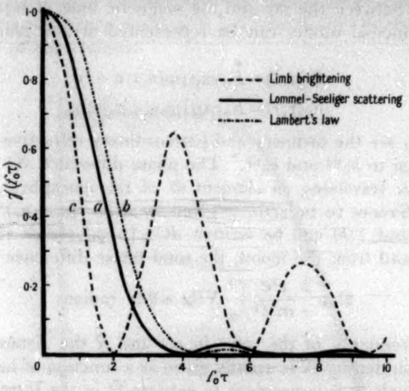


Figure 3. The pulse-to-pulse auto-correlation coefficient.

2.4. *The Effect of the Ionosphere on Moon Echoes*

The ionosphere has three principal effects upon radio waves passing through it: (*a*) absorption, (*b*) refraction of the beam with subsequent attenuation due to divergence and (*c*) changes of the state of polarization (magneto-ionic effect). At a frequency of 120 Mc/s, and for beam elevations between 40° and 65°, the two-way absorption through the ionosphere will always be less than 0·2 dB, while the refraction of the beam is always less than 0·1° (Mitra 1952) so that effects (*a*) and (*b*) may be neglected. On the other hand, as shown below, the magneto-ionic effect, for the frequency and beam directions used in the present experiments, results in a rotation of the plane of polarization of linearly polarized waves.

It can be shown that for ionospheric critical frequencies less than about 15 Mc/s, for wave frequencies of 120 Mc/s, and for ray directions within about 60° of the terrestrial magnetic vector, the quasi-longitudinal approximation to the Appleton–Hartree magneto-ionic formula can be used to calculate refractive indices to within 1% (Rawer 1953, p. 30). The principal modes of propagation are both circularly polarized, and the dielectric constants for them at any point are given by

$$k_0 = 1 - p_c^2/p(p + p_L) \quad \text{ordinary ray} \qquad \ldots\ldots(1\,a)$$

and

$$k_e = 1 - p_c^2/p(p - p_L) \quad \text{extraordinary ray} \qquad \ldots\ldots(1\,b)$$

where p is the angular frequency of the wave, p_c the angular ordinary ray critical frequency $=(4\pi\mathbf{c}^2N\mathbf{e}^2/\mathbf{m})^{1/2}$, p_L the longitudinal gyro-frequency $=\mathbf{e}B\cos\theta/\mathbf{m}$, N the number of electrons per cm^3, B the terrestrial magnetic induction (gauss) and θ the angle between the ray and the magnetic lines of force. The electric fields of the principal modes can be represented at any point along the ray direction $\mathbf{r}$ by

$$E_o(\mathbf{r},t)=E_o\exp[ip(n_or/\mathbf{c}-t)]$$
$$E_e(\mathbf{r},t)=E_e\exp[ip(n_er/\mathbf{c}-t)]$$

where n_o and n_e are the ordinary and extraordinary refractive indices, and are respectively equal to $k_o{}^{1/2}$ and $k_e{}^{1/2}$. The phase difference $d\Omega$ between the two components after traversing an element dr of the ionosphere, assuming their initial phase difference to be zero, is given by $d\Omega=p(n_o-n_e)/\mathbf{c}\,dr$ which from equations 1(a) and 1(b) can be written $d\Omega\simeq[p_Lp_c{}^2/p^2\mathbf{c}]dr$ radians. For the double path to and from the moon, the total phase difference is therefore

$$\Omega=\frac{2}{\pi}\frac{\mathbf{e}^3\mathbf{c}}{\mathbf{m}^2f^2}\int_0^R NB\cos\theta\,dr \text{ radians}$$

where f is the frequency of the wave in c/s and R the distance of the moon. Since the electron density N is usually given as a function of height h above the surface of the earth, it is convenient to express Ω in the form

$$\Omega=\frac{2}{\pi}\frac{\mathbf{e}^3\mathbf{c}}{\mathbf{m}^2f^2}\int_0^R NB\cos\theta\sec\chi\,dh \text{ radians} \qquad \ldots\ldots(2a)$$

where χ is the angle made by the ray to the zenith at height h. Putting in numerical values for constants (including $f=120$ Mc/s), this reduces to

$$\Omega=6{\cdot}71\times10^{-12}\int_0^R NB\cos\theta\sec\chi\,dh \text{ radians.} \qquad \ldots\ldots(2b)$$

For $\Omega=2n\pi$ ($n=0, 1, 2, 3\ldots$) the two circularly polarized waves are equivalent to a linearly polarized wave of the same plane of polarization as that emitted, while for $\Omega=(2n+1)\pi$, the resultant wave will have its plane of polarization perpendicular to that of the emitted wave. In general, the amplitude of the echo can be expressed by $E=E_0\cos(\Omega/2)$ where E_0 is the maximum amplitude of the electric vector. If approximate values are inserted for the quantities in the integrand ($B=0{\cdot}5$ gauss, $\theta=0°$, $\chi=45°$, $N=2\times10^5$ electrons cm^{-3}, depth of ionosphere $=150$ km), it can be seen that the total phase shift will be of the order of 20 radians, so that the overall changes of electron density of 10% will produce large changes in the mean amplitude of the echo. Experimental evidence for this effect is given in §4.3.

In order to make full use of the formula (2b), numerical values of the quantity $B\sec\chi\cos\theta$ were computed for several altitudes of the moon at transit. The curvature of the earth was taken into account, and a simple dipole magnetic field was assumed with the following parameters:

latitude of the boreal pole = 79° N.
longitude of the boreal pole = 70° W.
latitude of Jodrell Bank = 53° N.
longitude of Jodrell Bank = 2° W.
total magnetic induction at
Jodrell Bank at ground level = 0·475 gauss.

The results of the calculation are plotted in figure 4 against height. From these curves it is easy to compute Ω for a given model of the ionosphere. The geographic latitude of any point along the path of a ray at a height of 500 km decreases by 4° for a lunar altitude of 45° at transit, and by 2·2° for an altitude of 65°, so that it is sufficiently accurate to use a model representing the ionosphere above the place of observation.

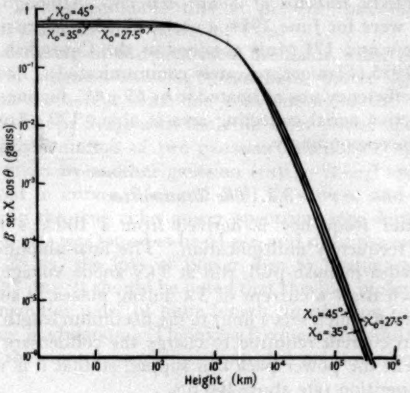


Figure 4. The variation of the function B sec χ cos θ with height.

§ 3. Apparatus and Calibration

The parameters of the equipment at present in use at Jodrell Bank are: frequency = 120 Mc/s, transmitter power = 10 kw peak (3 kw before January 1955), transmitter pulse length = 30 msec pulse repetition frequency = 0·6 p/s, receiver noise factor = 5 dB (8 dB before January 1955), energy bandwidth of receiver = 26 c/s and physical area of aerial = 250 m².

3.1. *The Aerial*

The aerial, which is used both for transmission and reception, is an echelon of ten elements arranged one behind the other along a North–South baseline. Each element consists of a reflecting screen tilted back at 45°, and illuminated by a double row of full-wave dipoles, giving a horizontally polarized electric vector. The length of each element transverse to the baseline is 4 wavelengths, while the width is 1 wavelength. The beam width between half-power points is 12° in azimuth and 5° in elevation. The beam is directed due South, and can be set to any elevation between 30° and 65°. The period of observation is limited by the aerial to 30 minutes before and after transit, for about a fortnight per month. The change of elevation of the beam is accomplished electrically by phase shifters, which are adjustable delays in the transmission lines feeding each element of the aerial. The theoretical form of the aerial directional diagram was confirmed by observations of the radio noise from the sun. These observations showed that the calculated and observed variations with azimuth of aerial gain agreed to within 3%.

The efficiency η of the aerial and its associated feeders can be expressed as the ratio of the observed to the theoretical power gain assuming no losses due to misphasing, etc. The efficiency was derived from measurements of the solar radio noise as the sun passed through the beam on 3rd–7th August 1955, using a noise diode as a comparison source. Reliable absolute measurements of the solar flux at 120 Mc/s are not available, but an estimate of $1{\cdot}6 \times 10^{-22}$ w m^{-2} (c/s)$^{-1}$ was obtained by interpolation from measurements at other frequencies given in the *I.A.U. Quarterly Bulletin of Solar Activity*. Although the most recent values published were for June 1954, a useful check was provided by the values of flux on 81 Mc/s and 175 Mc/s obtained at the Cavendish Laboratory from 3rd–7th August 1955 (Elsmore, privately communicated). From these measurements, the aerial efficiency was estimated to be 67 ± 4% for one-way transmission. That is, the effective aerial collecting area is about 170 ± 7 m^2, and the power gain is reduced correspondingly.

3.2. *The Transmitter*

The transmitter frequency is derived from a 100 kc/s master crystal by several stages of frequency multiplication. The final amplifier contains a pair of QY4–250 tetrodes in push-pull, run at 9 kv anode voltage. The power for these valves, which draw a current of 3 A during pulses, is supplied by a 50 μF condenser, the size of which sets a limit to the maximum length of the transmitter pulse. The mean current required to charge the condensers is 50 mA; this is the greatest current the power pack can supply, so that it is not permissible to raise the pulse repetition rate above 0·6 p/s.

The peak transmitter power is measured by using the transmitter output to heat a matched strip lamp. The mains power needed to heat the lamp to the same brightness is then determined by using a photometer. For these measurements the pulse repetition frequency is raised to 150 p/s and the pulse duration shortened to 100 μsec.

3.3. *The Receiver*

A cascode radio-frequency head is used, followed by amplification at five different intermediate frequencies, the last of which is 1 kc/s. The tuning of the receiver must be maintained at the correct frequency to within about 5 c/s over a period of several minutes. This is accomplished by deriving the frequencies of the first three local oscillators from the 100 kc/s crystal used in the transmitter. The fifth local oscillator frequency may be varied by ± 100 c/s to compensate for the Doppler shift in the frequency of the moon echoes (§ 2.2). The final intermediate-frequency amplifier has an energy bandwidth of 26 c/s. The detector is linear for signals of amplitude up to 100 times noise.

3.4. *Recording Technique*

A 12-inch long-afterglow range–amplitude display is used for visual monitoring while the part of the time-base which contains the echo is selected by a strobe, and photographed on continuously moving film. In addition, a loudspeaker fed from the undetected output of the last intermediate-frequency amplifier is used for aural monitoring.

3.5. *Signal Losses due to Frequency Modulation*

In constructing the equipment there was great difficulty in eliminating unwanted frequency modulation which appeared in the signals from the local oscillators and particularly in the signals from the transmitter. The frequency modulation, which appears to originate in small phase changes in the low frequency stages of the frequency multipliers, widens the frequency spectrum of the echo with consequent loss of echo power in the narrow-band receiver. Using the following method, it was estimated that at present the residual frequency modulation leads to a loss in echo power of 4 dB.

By changing the tuning of the final local oscillator while echoes are being received, a curve is obtained showing mean echo power against frequency (figure 5 (*b*)). This curve represents the convolution of the receiver power-frequency response (figure 5 (*a*)) with the true echo power spectrum. It can be shown that the convolution of two gaussian functions ($\exp[-(x/\sigma_1)^2]$ and $\exp[-(x/\sigma_2)^2]$) is given by another gaussian ($\exp[-\{x/(\sigma_1^2+\sigma_2^2)^{1/2}\}^2]$). Thus gaussians were fitted to curves 5 (*a*) and 5 (*b*) as shown, and from these the gaussian representing the true echo power spectrum was derived. The total energy bandwidth of this was found to be 64 c/s, and since the energy bandwidth of the receiver is 26 c/s, the loss of power due to frequency modulation is estimated as $3{\cdot}9 \pm 0{\cdot}5$ dB. It should be noted that this loss includes that inherent in the use of a transmitter pulse of finite length, in addition to that due to the unwanted frequency modulation. The extra frequency modulation introduced by the libration of the moon is negligible from this point of view.

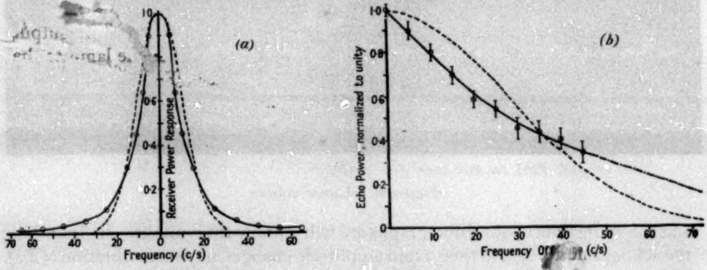


Figure 5. (*a*) The receiver power response, shown together with the fitted gaussian curve (broken). (*b*) The variation of echo power observed when the receiver is detuned together with the fitted gaussian curve (broken).

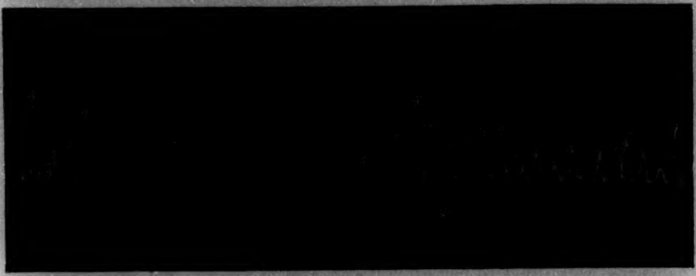
§ 4. RESULTS OF OBSERVATIONS

Observations of moon echoes have been made at frequent intervals from September 1953 to August 1955. During this period many improvements have been made to the equipment, so that it has not been possible to make observations regularly each month.

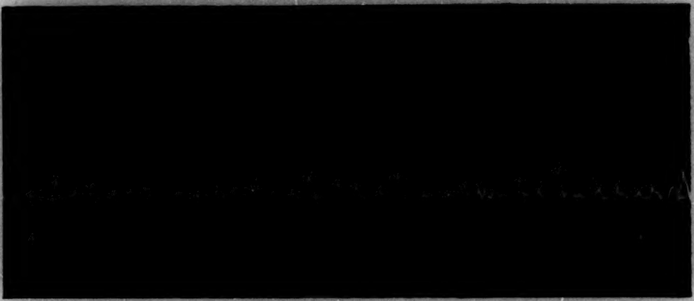
4.1. *The Echo Intensity*

Reliable measurements of echo intensity are made difficult by the fast fading caused by libration, and by the slow fading due to the magneto-ionic effect.

These are illustrated in figure 6. Each echo is obtained by photographing that part of the A-scan display where it appears. Thus the photographs 6 (*a* and *b*) are samples of successive echoes. The time interval between each echo is 1·8 sec and the white dots on the photographs are half-minute time-marks. In figure 6 (*a*) the echoes are strong and their rapid amplitude changes are due to libration (§ 2.3). The echoes in 6 (*b*) were obtained only 14 minutes before those in 6 (*a*). The aerial gain was substantially constant over this period. The rapid fading is still present, but the mean amplitude of the echoes is considerably less due to the slow fading produced by the magneto-ionic effect (§ 2.4).



(*a*)



(*b*)

Figure 6. Lunar echoes.

The effects of fast fading can be reduced by averaging intensities over a number of successive echo pulses. Since the intensities follow a Rayleigh law, the standard deviation of a single measurement is equal to the mean intensity I_0, so that when the average intensity of N pulses is measured, the standard deviation of the mean is given by $I_0/\sqrt{N}$. Averages over times greater than about two minutes are unreliable, however, because of the slow fading, so that with a pulse repetition frequency of about 0·6 p/s, the mean echo intensity cannot be measured better than about 15%. The slow fading introduces a further difficulty that, in order to measure the maximum possible intensity, it is necessary to observe at a time when the echo is horizontally polarized.

In order to test the relation between the observed and predicted echo intensities, special observations were made on 10th and 16th August 1955, these dates being chosen so that the elevation of the moon (54°) would be the same as that of the sun for the dates on which the solar noise was used to calibrate the aerial. All other parameters of the equipment were carefully checked on these days, the losses due to the aerial and to frequency modulation being taken into account (cf. §§ 3.2, 3.5). The following results were obtained:

	10.8.55	16.8.55
Predicted signal-to-noise ratio (dB)	31·8 ± 1	31·0 ± 1
Max. mean observed signal-to-noise ratio averaged over one minute (dB)	27·2 ± 2	24·3 ± 2

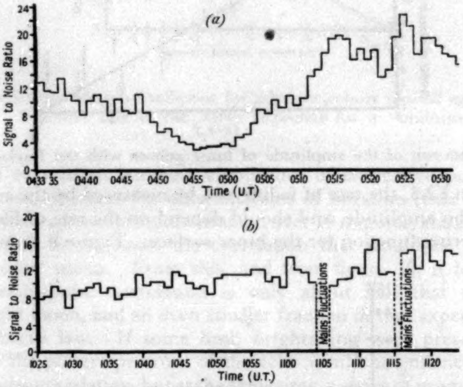


Figure 7. (*a*) The signal-to-noise ratio for the lunar run on 10th August 1955.
(*b*) The signal-to-noise ratio for the lunar run on 16th August 1955.

Figures 7 (*a*) and (*b*) show for each run the variation with time of the mean echo intensity averaged over one minute. It can be seen that on 16th August there was no clearly defined maximum in the signal intensity, so that the observed echo intensity for this date is probably a lower limit.

On the assumption that the directivity of the moon is 5·7 and that the power reflection coefficient of its surface is 0·1, the observed echo intensity is lower than that predicted by 5 ± 3 dB. It should be noted that the discrepancy is considerably less than that found by other workers (cf. § 1).

4.2. *Rapid Fading of Moon Echoes*

If the moon can be regarded as an assembly of random scatterers, the number of echoes of amplitude greater than z is given by the Rayleigh law (§ 2.2):

$$N(z) = N_0 \exp[-(z/z_0)^2]$$

where z_0 is the r.m.s. amplitude and N_0 the total number of echoes observed. Figure 8 shows a plot of $\log N(z)$ against $(z/z_0)^2$ for two five-minute periods. It can be seen that the fit to the Rayleigh law is excellent. The small departures from the Rayleigh law for large amplitudes are partly due to the effects of the long-period fading producing changes in the value of z_0. It should be noted that it is not possible from these measurements to deduce the total number of scatterers, because a close approximation to the Rayleigh law is given by any number of scatterers greater than about 10 (Lawson and Uhlenbeck 1950, p. 47).

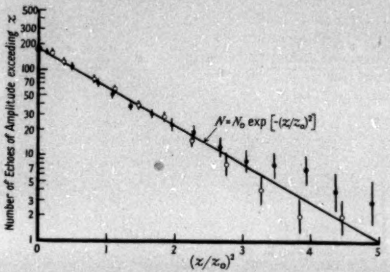


Figure 8. Comparison of the amplitude of lunar echoes with the Rayleigh distribution.

As shown in § 2.3, the rate of fading can be measured by the auto-correlation function for echo amplitude, and should depend on the rate of libration, as well as on the scattering function for the lunar surface. Figure 9 shows a plot of the

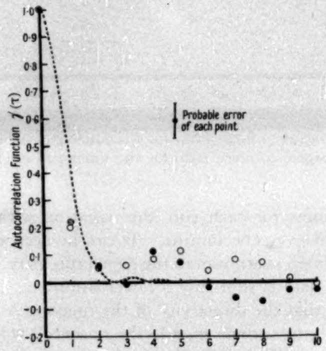


Figure 9. The autocorrelation function for two samples of 300 echoes. The broken curve is that expected for a 'uniformly bright' lunar disc.

observed auto-correlation function, corrected for the effects of receiver noise (Lawson and Uhlenbeck 1950, p. 134) for the two five-minute periods represented in figure 8. This diagram also shows the expected auto-correlation function for a uniformly 'bright' moon, when the appropriate rate of libration for the date is used. It can be seen that the observed auto-correlation function falls quickly to zero in a time comparable with the pulse interval, and thereafter shows only random oscillations about zero. These are almost certainly due to errors of measurement and to the finite size of the sample. Figure 10 shows mean values for the correlation between amplitudes of successive pulses, corrected for receiver noise, obtained for different rates of libration (that is, on different days). Each point represents a measurement over 10 minutes, about 300 successive pulses. The full curve shows the correlation that would be expected for a uniformly 'bright' moon. From this, and from figure 10, it is clear that the observed pulse-to-pulse correlation is only about half that expected for a uniformly bright moon, and an even smaller fraction of that expected for a moon obeying Lambert's law. If some limb brightening were present, the higher frequencies in the power spectrum of the echo would be enhanced, so reducing the pulse-to-pulse correlation, but at the same time, a series of prominent secondary maxima would appear in the auto-correlation function (cf. figure 3). No such maxima are indicated by the observations, so that it is not possible to explain the smallness of the observed pulse-to-pulse correlation by a measure of limb 'brightening'. In view of this, it becomes necessary to suppose that some mechanism is destroying the correlation in the signals between the transmitter and the receiver. Such a mechanism may be provided by irregularities in the ionosphere—rapid changes in the total electron content along the line-of-sight will produce corresponding oscillations in the plane of polarization of the echo, and with them, changes in the echo intensity. To produce the observed loss in correlation, the phase angle Ω (§2.4) would have to oscillate by about 20° with a period of the order of a few seconds. Such changes would be given by ionospheric irregularities which changed the total electron content in a column of 1 cm^2 cross section along the line-of-sight by about 5×10^{10} electrons. Such an irregularity would be given, for example, by the ionized trail of a meteor brighter than about 3rd magnitude, which would persist for a few seconds. Another possible cause of loss of correlation between pulses is phase instability in the transmitter (§3.5).

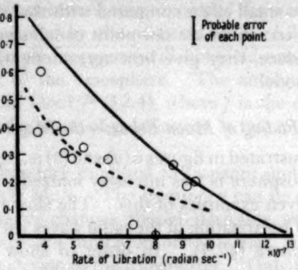


Figure 10. The autocorrelation coefficient for adjacent echoes plotted against the moon's libration. The full line is the curve expected for a 'uniformly bright' lunar disc.

To summarize this section, it may be said that, while the observed auto-correlation function for echo amplitude shows no detectable secondary maxima indicative of limb-brightening, yet the correlation is less than would be expected for a uniformly bright moon by 40% and for a limb-darkened moon (Lambert's law) by 55%. Since the observations show a systematic change of fading rate with the rate of libration, any additional fading mechanism such as that postulated above must give only a small effect compared with that due to the libration itself. Thus although the observations do not point conclusively to any one scattering law at the moon's surface, they give best agreement with the uniformly bright (Lommel–Seeliger) model.

4.3. *Slow Fading of Moon Echoes—the Magneto-ionic Effect*

The slow fading illustrated in figures 6 (*a*) and (*b*) is, in general, less pronounced at night when the ionosphere is less intensely ionized; Murray and Hargreaves (1954) have already given examples of this. The slow fading is attributed to the rotation of the plane of polarization of the radio waves with changing total electron content in the ionosphere (§ 2.4). In order to show the correctness of this explanation, an experiment was carried out during March 1954 using for reception a paraboloid aerial 10 metres in diameter, with both horizontally and vertically polarized dipoles at its focus. These were connected to separate preamplifiers, and the main receiver was switched from one to the other every minute. The signals were weak, and hard to measure accurately, because the power gain of the paraboloid is some 8 dB less than that of the echelon array, while at the same time the overall gain of the equipment was some 10 dB less than at present. It was found that while the transmitted signal was horizontally polarized, the received echoes were often more intense on the vertically polarized dipole. Figure 11 shows a plot of the variation of signal strength observed in the two planes of polarization, suitably corrected for the polar diagrams of the aerials, together with the sum of the intensities observed on the two aerials. If slow fading is due to the rotation of the plane of polarization of the wave, the sum of the intensities in any two perpendicular planes of polarization should be constant. It can be seen that this quantity is roughly constant, the apparent increase being attributed to drifts in receiver gain. It is interesting to note that, in figure 7 (*a*), the signal intensity does not fall to zero at the minimum, an indication that the radio waves are

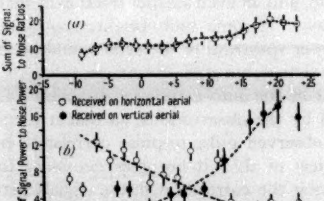


Figure 11. (*a*) The total received power using mutually perpendicular aerials in March 1954; the slight rise is probably due to drifts in the equipment. (*b*) The signal-to-noise ratio on each aerial during the run shown above.

undergoing some 10% depolarization, probably by secondary scattering at the surface of the moon.

From a single measurement of the angle to the horizontal made by the plane of polarization of the echo, it is not possible to determine the total phase difference Ω between the two circularly polarized components of the wave, because the plane of polarization will be unaltered by the addition to Ω of $2\pi n$ $(n=0,1,2,3\ldots)$. Thus measurements at a single frequency will not yield unambiguous values for the total electron content of the ionosphere. The ambiguity can be resolved with the help of the relation $\Omega \propto 1/f^2$ (§ 2.4), where f is the radio frequency, and by measuring the angle of the plane of polarization at two or more different radio frequencies.

The equipment was therefore modified to transmit and receive frequencies of 119·28, 120·00 and 120·72 Mc/s. The frequency can be changed by the turn of a switch. Larger frequency changes, though desirable, are not easy to obtain, partly because of the bandwidth limitation of the aerial, and partly because of limitations imposed by the design of the transmitter. Only preliminary observations, using the horizontally polarized aerial array, have so far been made.

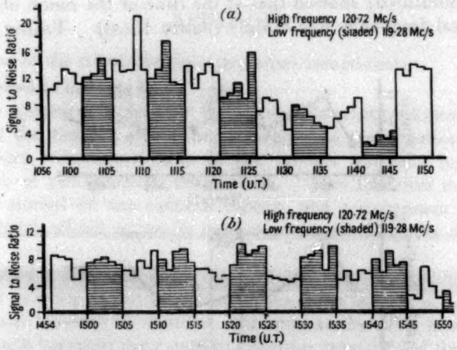


Figure 12. (*a*) The signal-to-noise ratio for the lunar run on 19th June 1955, showing a fall from a maximum into a minimum on both frequencies, and the displacement of the minima. (*b*) The signal-to-noise ratio for the lunar run on 23rd June 1955. Here the run does not contain sufficient information to enable the analysis described in the text to be applied.

Figure 12 shows sample runs during which the frequency was switched between 119·28 and 120·72Mc/s every five minutes. The phase angle will be changed by about 2·5% by this shift in frequency.

Records such as those in figure 12 are analysed on the assumption that, during any one run, the rate of change of total ionospheric electron content is constant. Then if the time taken for the echo intensity to fall from a maximum to a minimum on the higher frequency is T, and the time difference between the occurrence of minima on high and low frequencies is t, reckoned positive when the minimum on the lower frequency follows that on the higher frequency, and if $\Omega_{\mathrm{L}}, \Omega_{\mathrm{H}}$ denote the values of the phase angle on the lower and higher frequencies respectively, then at the instant when the signal intensity is a minimum on the lower frequency, we have:

$$\Omega_L = (2n+1)\pi \qquad (n = 0, 1, 2 \ldots),$$
$$\Omega_H = [2(n-q)+1]\pi \pm \pi t/T.$$

The two frequencies are $f\sqrt{(1-\delta)}$, f, with δ positive, and the upper and lower signs signify that Ω is respectively increasing or decreasing with time, and q is any integer or zero.

But

$$f^2\Omega_H = f^2(1-\delta)$$
$$(2n+1)\pi(1-\delta) = [2(n-q)+1]\pi \pm \pi t/T$$

whence

$$\Omega_H = (2q \mp t/T)/\delta \text{ radians.}$$

Thus on 19th June 1955, $t = 8$ min, $T = 25$ min, and so at 1140 U.T. we find $\Omega_H = 40\pi/3$ radians ($q = 0$, with total electron content decreasing) or $\Omega_H = 200\pi/3$ radians ($q = 1$, with the total electron content increasing). Larger values of Ω_H corresponding to higher values of q are also possible. Ionosphere soundings at Slough (Smith-Rose, privately communicated) showed that at the time of the moon observations, the F2 region critical frequency was 6 Mc/s (figure 13(*a*)). Taking this value, the maximum value of the total electron content along the line-of-sight cannot have exceeded $2{\cdot}5 \times 10^{13}$ electrons cm^{-2}, even if the maximum electron density in the F2 region is assumed to extend over as much as 500 km in height. This value of total electron content would give a phase angle Ω_H of about 100 radians. Thus, of the various possibilities, the observed value of $\Omega_H = 40\pi/3$ radians seems the most reasonable to adopt. This value implies, however, that during the observation (1050–1150 U.T.) the total electron content of the ionosphere decreased by some 15%—a result in rough agreement with the Slough data for the product $(f_0F2)^2y_m$ where y_m is the half-thickness of the F2 region, assuming this to be

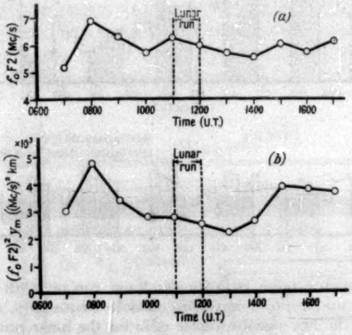


Figure 13. (*a*) The variation of f_0F2 on 19th June 1955. The period of the lunar run is indicated. (*b*) The variation cf $(f_0F2)^2y_m$ on 19th June 1955. During the period of the lunar run the electron content $(f_0F2)^2y_m$ fell by 10%.

a measure of the extent of the ionosphere (figure 13 (*b*)), since for a parabolic layer the electron content up to the maximum f_0F2 is $8{\cdot}3 \times 10^{-9}(f_0\mathrm{F2})^2 y_\mathrm{m}$.

It is not proposed here to discuss further the interpretation of this result, or its relation to the question of the distribution with height of the ionization in the F region (Ratcliffe 1951, Chatterjee 1954). More detailed results and deductions from them, together with a description of the switched-frequency equipment, will be presented after an extended series of observations.

ACKNOWLEDGMENTS

We are indebted to Professor A. C. B. Lovell for his encouragement and for providing the facilities at Jodrell Bank, to Dr. J. G. Davies for valuable help in the design of the equipment, and to Dr. T. R. Kaiser for his most generous help and advice. We also wish to thank the Director of the Radio Research Station, Slough, for providing us with pre-publication ionosphere records.

APPENDIX

The libration of the moon has four principal components:

1. *The libration in latitude* l_β.

The moon's equator is inclined to the plane of its orbit by about 6·5°, and since the direction of the moon's axis is fixed (neglecting perturbations) an observer on the Earth sees sometimes more of the moon's northern hemisphere, and sometimes more of the southern hemisphere. This libration in latitude, as it is called, has a period of one sidereal month, and a maximum value of about 3×10^{-7} radian sec^{-1}, which occurs as the moon crosses the nodes of its orbit.

2. *The libration in longitude* l_L.

The angular velocity of the moon's rotation about its axis is equal to its *mean* geocentric angular velocity in its orbit, but because the orbit is an ellipse, there will be a difference between the rotational angular velocity and the instantaneous geocentric orbital angular velocity. This leads to a libration in longitude, which is direct in sense at apogee, and retrograde at perigee, with a maximum value of about 4×10^{-7} radian sec^{-1}.

3. *The diurnal libration* l_D.

As the earth rotates, an observer on its surface sees a parallactic shift between the centre of the moon's projected disc and the true centre of the moon. This so-called diurnal libration is always in a direct sense when the moon is in transit, with a maximum value for declination 0° of about $12 \cos \Psi \times 10^{-7}$ radian sec^{-1}, where Ψ is the geographic latitude of the observer.

4. *Physical libration.*

This is a rocking produced by the variable gravitational couples exerted by the earth and sun on the moon's equatorial bulge. Its maximum value of about 3×10^{-10} radian sec^{-1} is so small in comparison with the other librations that it may be neglected.

The resultant value of the libration is obtained by vector addition of the components l_B, l_L and l_D (neglecting the physical libration). If a system of cartesian coordinates be set up with x axis parallel to the moon's radius vector, libration in longitude l_L may be represented by a vector normal to the moon's orbit (z axis), libration in latitude l_β by a vector parallel to the y axis, and diurnal libration l_D, by a third vector parallel to the axis of the earth. During a sidereal month, vector l_D precesses around the z axis making with it a constant angle δ_0 equal to the greatest declination reached by the moon during that month. The projection of l_D into the xy plane makes with the y axis an angle Θ approximately equal at any instant to the difference between the longitude of the moon and the mean longitude of the ascending node. The component of the resultant libration along the x axis is not of course a true libration and introduces no phase changes into the paths of rays to different points on the moon's surface. Thus the effective total libration l_T at transit is given by:

$$l_T^2 = (l_L + l_D \cos \delta_0)^2 + (l_\beta + l_D \sin \delta_0 \cos \Theta)^2.$$

The quantities l_L, l_β, δ_0, Θ can be readily found from the *Nautical Almanac*.

REFERENCES

BAY, Z., 1946, *Hung. Phys. Acta*, **1**, 1 (quoted by Kerr and Shain 1951).

BOOKER, H. G., RATCLIFFE, J. A., and SHINN, D. H., 1950, *Phil. Trans. Roy. Soc.* A, **242**, 579.

CHATTERJEE, B., 1954, *Nature, Lond.*, **173**, 263.

DE WITT, J. H., and STODOLA, E. K., 1949, *Proc. Inst. Radio Engrs, N.Y.*, **37**, 229.

GRIEG, D. D., METZGER, S., and WAER, R., 1948, *Proc. Inst. Radio Engrs, N.Y.*, **36**, 652.

KERR, F. J., and SHAIN, C. A., 1951, *Proc. Inst. Radio Engrs, N.Y.*, **39**, 230.

KERR, F. J., SHAIN, C. A., and HIGGINS, C. S., 1949, *Nature, Lond.*, **163**, 310.

LAWSON, J. L., and UHLENBECK, G. E., 1950, *Threshold Signals* (New York : McGraw-Hill).

MITRA, A. P., 1952, *Indian J. Phys.*, **26**, 495.

MURRAY, W. A. S., and HARGREAVES, J. K., 1954, *Nature, Lond.*, **173**, 944.

NORTON, K. A., and OMBERG, A. C., 1947, *Proc. Inst. Radio Engrs, N.Y.*, **35**, 4.

PIDDINGTON, J. H., and MINNETT, H. C., 1949, *Aust. J. Sci. Res.* A, **2**, 63.

RATCLIFFE, J. A., 1951, *J. Geophys. Res.*, **56**, 487.

RAWER, K., 1953, *Die Ionosphäre* (Groningen : Noordhoff).

RUSSELL, H. N., DUGAN, R. S., and STEWART, J. Q., 1945, *Astronomy* : I-*The Solar System* (Boston : Ginn).

SULZER, P. G., MONTGOMERY, G. F., and GERKS, I. H., 1952, *Proc. Inst. Radio Engrs, N.Y.*, **40**, 361.

A General Theory of the Surface Friction of Solids

By C. RUBENSTEIN
The British Cotton Industry Research Association, Shirley Institute, Didsbury,
Manchester 20

*Communicated by F. C. Toy; MS. received 12th December 1955,
and in revised form 23rd February 1956*

Abstract. The theory of the surface friction of solids which is developed is based on the assumptions that the adhesion theory of friction is valid and that the asperities, at which contact occurs, deform according to the law stress$\propto$(strain)^x. A roughness model is postulated in which the asperities (either spherical or cylindrical) (*a*) are uniformly distributed over the surface, (*b*) have the same radius of curvature and (*c*) all protrude from the surface by the same amount. The analysis leads to the equation $F=CL^\beta$ relating the frictional force F to the normal load L, an equation which has been found empirically by several investigators. The parameter C is shown to be dependent on the physical properties of the sliding materials, on the degree of surface roughness and on the apparent area of contact, and hence, on the geometrical arrangement of the sliding surfaces, while the parameter β is dependent on the physical properties of the sliding materials only. Amontons' laws of friction (the particular case of the above equation when β is unity) are shown to be interdependent, and when these are obeyed the frictional force is shown to be independent of the degree of surface roughness provided this is not excessive. The influence of surface roughness on the frictional force, found experimentally, is shown to be in accordance with deductions based on the adhesion theory of friction and opposed to deductions based on the Coulomb theory of friction.

§ 1. INTRODUCTION

THE typical frictional behaviour of non-metals is found to differ from that of metals in three important respects. Thus, numerous investigations (see below) have revealed that, in general, the coefficient of friction μ of a non-metal is dependent on load, increasing as the load is decreased. The coefficient of friction of metals, on the other hand, tends in general to be independent of the normal load (Amontons' first law) over a wide range of loads (Whitehead 1950). Secondly, the friction of non-metals is dependent on surface roughness (e.g. Roth, Driscoll and Holt 1942, King 1950, Lincoln 1952, Denny 1953), the frictional force decreasing with increase in roughness, while the friction of metals appears to be independent of the degree of surface roughness, provided this is not excessive (Strang and Lewis 1949, Bowden and Tabor 1950). Finally, the friction of non-metals is found to be dependent on the apparent area of contact, the frictional force increasing as the apparent area of contact increases (e.g. Howell 1951, Lincoln 1952, Pascoe and Tabor 1955) while the friction of metals is in general independent of the apparent area of contact (Amontons' second law).

Previous attempts to obtain a theoretical explanation of frictional phenomena have been aimed at explaining specifically the behaviour either of metals (Holm 1938, Merchant 1940, McFarlane and Tabor 1950) or of non-metals (see below). It will be shown that an explanation of the typical frictional behaviour of metals can be derived as a limiting case of a more general theory of the friction of solids, and that while the more general behaviour is exhibited by many non-metals over a wide load range, it is exhibited by all but the softest metals at very low loads only. Consequently, we begin by stating the assumptions made in previous theories of the friction of non-metals. For convenience, these theories are grouped as follows: (1) wherein the material is assumed to be Hookean and contact is assumed to occur at numerous points within the apparent area of contact (Lincoln 1952, 1953, Archard 1953, Lodge and Howell 1954); (2) wherein the material is assumed to have a well-defined yield stress and the specimens are assumed smooth, the area of contact consisting of a central zone produced by plastic deformation and an outer zone produced by elastic deformation (Gralén, Oloffsen and Lindberg 1953); (3) wherein the deformation law of the material is obtained empirically from static loading experiments and the specimens are assumed smooth (Pascoe and Tabor 1955).

In the present work the assumptions made are more general than those of (1) or (3) above, the results of these latter analyses appearing as special cases of the relationships derived in the next section. In subsequent sections deductions from the present theory will be compared with experimental observation, when the extent of agreement may be assessed.

§ 2. THEORETICAL CONSIDERATIONS

Tabor (1948) has shown that the data obtained from hardness measurements of a metal indented by a hard sphere can be correlated with the data obtained by stressing the metal in uniaxial tension or in frictionless compression if $\frac{1}{3}$(Meyer hardness), i.e. the load divided by the projected area of the indentation, is plotted as the stress and $\frac{1}{5}(d/D)$ is plotted as the strain, where d is the diameter of the indentation and D is the diameter of the indenter. Similar correlation is obtained if the geometrical arrangement is reversed, i.e. if a metal hemisphere is loaded against a hard plane surface. In what follows, it will be assumed (assumption (i)) that for non-metals, the stress can be represented as $\theta \times$ the Meyer hardness and the strain by $\phi(d/D)$, where θ and ϕ are numerical constants.

Any continuous curve can be fitted by a power-law relationship over a limited range and, in particular, the stress–strain curve of a non-metal can be fitted by the equation†

$$\sigma = b\epsilon^{x} \qquad \ldots\ldots(1)$$

where σ is the uniaxial stress and ϵ is the natural strain, with b and x constant over a limited strain range and varying, if necessary, from one strain range to the next. As will be shown below, the friction and indentation data indicate that over a very wide range of strain the deformation of some non-metals can be represented by equation (1) with single values of b and x. Many materials (e.g. metals and crystalline non-metals) tend to be elastic at low strains and plastic at high strains and equation (1) is particularly well suited to represent this gradation, for, from equation (1), when $x=1$ the stress is proportional to the strain, the condition of elasticity, while, when $x=0$, the stress is constant, the condition of plasticity.

---

† This equation has been shown to represent the stress–strain curve of some metals over a wide range of strain (Nadai 1931).

Applying equation (1) to the case of contact between a hemisphere and a plane (after Tabor 1948)

$$\frac{L}{d^2} = \frac{\pi}{4} b \frac{\phi^x}{\theta} \left(\frac{d}{D}\right)^x$$

i.e.

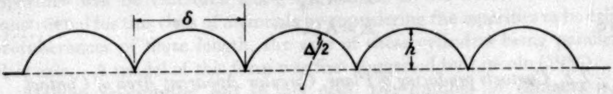
$$L = K_1(d^{x+2}/D^x) \qquad \ldots\ldots(2)$$

where $K_1 = \frac{1}{4}\pi b\phi^x/\theta$. The area of contact

$$A = \tfrac{1}{4}\pi d^2 = \tfrac{1}{4}\pi(LD^x/K_1)^\alpha \qquad \ldots\ldots(3)$$

where $\alpha = 2/(2+x)$. Equations (2) and (3) have been derived for the particular case of contact between a plane surface and a specimen of spherical shape. They are, however, of more general application and apply to any system in which the area of contact is circular. Hence they apply, with a different value of K_1, to the contact of two cylinders of equal diameter D whose axes are at right angles to each other, and also to the contact between two hemispheres of equal diameter. In the case of two cylinders of equal diameter, inclined to each other at an angle less than 90°, the apparent area of contact is elliptical. The major and minor axes of the ellipse, d_1 and d_2, however, are both related to the load by an equation of the form of equation (2), the values of K_1 being different for each axis. Hence the area of contact, which is proportional to the product of d_1 and d_2, is given by equation (3) with a value of K_1 different from that applicable to the deformation of a sphere.

It will now be assumed that the area of contact at an asperity and the apparent area of contact between the friction members both vary with the load according to deformation laws of the same form, provided the asperity contact is geometrically similar to the bulk contact (assumption (ii)). It follows from this assumption that equations (2) and (3) apply to the contact between a curved asperity and a hard plane surface or between two curved asperities where L now represents the load at an asperity (L_i), D the diameter of the asperity (Δ), d the diameter of the area of contact (d_i) and A the area of contact at an asperity (A_i). In the following sections x_1 and α_1 will be used to represent the values of the parameters x and α at bulk contacts and x_2, α_2 at asperity contacts.



The following further assumptions are made: (iii) the frictional force is proportional to the true area of contact; the constant of proportionality can be regarded as a shear strength S_0; (iv) the shear strength S_0 at the points of contact is constant, independent of the normal pressure or the strain when slip occurs; (v) the normal and tangential forces are independent of each other. [Assumptions (iv) and (v), while arbitrary, are commonly made and lead to relationships which offer a satisfactory qualitative explanation of the frictional behaviour of most materials.] (vi) As an idealized concept of roughness, all the asperities will be assumed to have the same radius of curvature $\Delta/2$, to protrude from the surface by the same amount h, as in the figure, and to be uniformly distributed over the surface; it follows as a corollary to this assumption that for plane surfaces all asperities make the same contribution to the total normal reaction and to the total frictional force. (vii) When the apparent area of contact is of circular contour, the normal pressure will be assumed to vary in accordance with a modified form of the Hertz (1881) equation, viz.

$$p_A = E\{\tfrac{1}{4}d^2 - r^2\}^\tau \qquad \ldots\ldots(4)$$

where p_A is the normal pressure at a point distance r from the centre of the area of contact of diameter d, and E is a constant. For a Hookean material $\tau = \frac{1}{2}$.

2.1. *Contact between Plane Surfaces*

Consider contact between a hard smooth plane surface and a plane non-metal surface, the latter being covered with asperities in accordance with assumption (vi). It will be assumed that the asperities are segments of spheres.

Let there be n asperities in contact when the surfaces are loaded normally with a load L.

Then the load per asperity is $L_i = L/n$. Applying assumption (ii), the area of contact at each asperity is given by equation (3), i.e. $A_i = \frac{1}{4}\pi(L\Delta^{x_2}/nK_1)^{\alpha_2}$. The total true area $= nA_i = n^{1-\alpha_2}\frac{1}{4}\pi(L\Delta^{x_2}/K_1)^{\alpha_2}$. Making assumptions (iii)–(vi) $F = S_0 nA_i = S_0 n^{1-\alpha_2}\frac{1}{4}\pi(L\Delta^{x_2}/K_1)^{\alpha_2}$. Let there be n_0 asperities per unit of the apparent area of contact A_A then $n = n_0 A_A$, i.e.

$$F = S_0\tfrac{1}{4}\pi(n_0 A_A)^{1-\alpha_2}(L\Delta^{x_2}/K_1)^{\alpha_2}. \qquad \ldots\ldots(5)$$

It is now proposed to take into account the influence of roughness (see assumption (vi)). From the figure it can be seen that $h(\Delta - h) = \frac{1}{4}\delta^2$. Since the asperities are assumed to be uniformly distributed over the surface, n_0 varies inversely as δ^2, i.e. $n_0 = K_2/[h(\Delta - h)]$. Substituting in equation (5)

$$\begin{aligned} F &= S_0\tfrac{1}{4}\pi\{K_2 A_A/[h(\Delta - h)]\}^{1-\alpha_2}(L\Delta^{x_2}/K_1)^{\alpha_2} \\ &= K_3(A_A\omega)^{1-\alpha_2}L^{\alpha_2} \end{aligned} \qquad \ldots\ldots(6)$$

where $K_3 = \frac{1}{4}\pi S_0(K_2^{1-\alpha_2}/K_1^{\alpha_2})$ and $\omega = 1/[\rho(1-\rho)]$. ω is the roughness factor and the roughness $\rho = h/\Delta$. It should be noted that on the model postulated, the maximum value of ρ is $\frac{1}{2}$.

Thus, the frictional force is seen to be dependent on the roughness provided $\alpha_2 \neq 1$. For ω constant equation (6) becomes

$$F = KA_A^{1-\alpha_2}L^{\alpha_2}$$

where

$$K = K_3\omega^{1-\alpha_2}. \qquad \ldots\ldots(7)$$

2.2. *Contacts producing a Plane, Circular Apparent Area of Contact*

It has been mentioned above that a circular apparent area of contact is produced with contacts between (i) a hemispherical specimen and a plane surface, (ii) two-hemispherical specimens of equal diameter and (iii) two cylinders of equal diameter whose axes are mutually perpendicular.

Consider an annular element of the area of contact enclosed by circles of radii r and $r + dr$. If dr is considered to be sufficiently small for the variation of the mean normal pressure over the element of area to be ignored, then the frictional force due to this element is given by equation (6), i.e.

$$dF = K_3 A_A^{1-\alpha_2}\omega^{1-\alpha_2}L^{\alpha_2} = K_3\omega^{1-\alpha_2}A_A p_A^{\alpha_2}.$$

From assumption (vii)

$$p_A = E(\tfrac{1}{4}d^2 - r^2)^\tau. \qquad \ldots\ldots(4)$$

Hence

$$dF = K_3\omega^{1-\alpha_2}E^{\alpha_2}(\tfrac{1}{4}d^2 - r^2)^{\alpha_2\tau}2\pi r\,dr.$$

Integrating over the area of contact between the limits 0 and $d/2$

$$F = \{\pi K_3\omega^{1-\alpha_2}E^{\alpha_2}/(\alpha_2\tau + 1)\}(\tfrac{1}{2}d)^{2(\alpha_2\tau+1)}. \qquad \ldots\ldots(8)$$

From equation (4)

$$L = \int_0^{d/2} E(\tfrac{1}{4}d^2 - r^2)^{\tau}2\pi r\,dr = \{\pi E/(\tau+1)\}(\tfrac{1}{2}d)^{2(\tau+1)} \ldots\ldots(9)$$

But from equation (3) $L = K_1(d^{x_1+2}/D^{x_1})$. Equating these expressions for L:

(i) $x_1 = 2\tau$, i.e. $\alpha_1 = 1/(1+\tau)$,

(ii) $E = \{K_1/D^{x_1}\}\{(\tau+1)/\pi\}2^{2(\tau+1)} = K_4/D^{x_1}$.

Substituting for d from equation (9) into equation (8)

$$F = \{\alpha_1\pi K_3\omega^{1-\alpha_2}/(\alpha_1+\alpha_2-\alpha_1\alpha_2)\}\{D^{x_1}/K_4\}^{\alpha_1(1-\alpha_2)}\{L/\alpha_1\pi\}^{\alpha_1+\alpha_2-\alpha_1\alpha_2} = CL^{\beta} \qquad \ldots\ldots(10)$$

where $C = K_5D^{x_1\alpha_1(1-\alpha_2)} = K_5D^{2(1-\alpha_1)(1-\alpha_2)}$, $K_5 = (\alpha_1\pi)^{1-\beta}K_3\omega^{1-\alpha_2}/\beta(K_4)^{\alpha_1(1-\alpha_2)}$ and $\beta = \alpha_1 + \alpha_2 - \alpha_1\alpha_2$.

The following particular cases are of interest:

(i) $\alpha_1 = \alpha_2 = \alpha$, i.e. $x_1 = x_2 = x$. Equation (10) is obtained with

$$C = K_5D^{2(1-\alpha)^2}, \quad K_5 = (\alpha\pi)^{1-\beta}K_3\omega^{1-\alpha}/\beta(K_4)^{\alpha(1-\alpha)} \qquad \ldots\ldots(11)$$

$$\beta = \alpha(2-\alpha) \qquad \ldots\ldots(12)$$

(ii) $\alpha_2 = 1$, i.e. $x_2 = 0$. Equation (10) is obtained with

$$C = K_5 = K_3 = S_0\theta/b = S_0/p_0 \ (p_0 \text{ is the Meyer hardness})$$

and $\beta = 1$, i.e. Amontons' law is obeyed when $\alpha_2 = 1$.

2.3. *Contact between Crossed Cylinders*

Measurement of the friction of textile fibres are the most common examples of the crossed cylinder arrangement. From the mode of production of man-made textile fibres (i.e. by extrusion) it might be expected that for these materials the asperities will be extended along the length of the fibre. A model can be constructed for this class of materials by considering the asperities to be cylindrical protuberances of finite length, the axes of these cylinders being parallel to the fibre axis. A model of this form was first suggested by Lincoln (1953).

Making assumption (vi) in this case, the number of contacts is, again, proportional to the apparent area of contact. For two fibres of equal diameter D and inclined to each other at right angles then, as shown in § 2.2(i), equation (10) is obtained with

$$C = K_5'D^{2(1-\alpha)^2} \ \ldots\ldots(11a) \qquad \beta = \alpha(2-\alpha) \ \ldots\ldots(12a)$$

provided $\alpha_1 = \alpha_2 = \alpha$. If the axes of the fibres are inclined at an angle less than 90° the true and apparent areas of contact will be elliptical, but as shown above, the apparent area of contact will be proportional to L^{α} and, provided the major axis of the ellipse of contact at an asperity is less than the length of the contact λ, the area of contact at an asperity will be proportional to L^{α}. λ is the length of overlap of two asperities of finite length in contact.

When the major axis of the elliptical contact at an asperity becomes equal to λ, further increase in the load can produce an increase in the minor axis only, so that in this case assumption (ii) is inapplicable. In this event, as an approximation the area of contact at an asperity can be taken as rectangular, the lengths of the sides of this rectangle being λ and d_2. If there are $2n$ asperities in contact, then at any contact $L/n\lambda d_2 = b(\phi^x/\theta)(d_2/\Delta)^x = K_1'(d_2/\Delta)^x$ where d_2 is the minor axis of the ellipse and $K_1' = 4K_1/\pi$, i.e. $A_i = \lambda d_2 = \lambda^{1-\gamma}\{L\Delta^x/(nK_1')\}^\gamma$ where $\gamma = 1/(1+x)$ and $x_1 = x_2 = x$. Hence the total area of contact $nA_i = (n\lambda)^{1-\gamma}(L\Delta^x/K_1')^\gamma$, i.e. $F = S_0(n\lambda)^{1-\gamma}(L\Delta^x/K_1')^\gamma$. If $n = n_0 A_A$ then $F \propto A_A{}^{1-\gamma}L^\gamma$. For two fibres inclined at an angle ψ where $\frac{1}{2}\pi > \psi > 0$, $A_A \propto L^\alpha$, hence the friction–load behaviour is again expressed by equation (10) with

$$C = K_5'' D^{2(1-\alpha)(1-\gamma)} \quad \ldots\ldots(11\,b) \qquad \text{and} \qquad \beta = \alpha(1-\gamma)+\gamma. \quad \ldots\ldots(12\,b)$$

For two parallel fibres, $\psi = 0$, assumption (ii) is valid, i.e. $A_A \propto L^\gamma$ and in this case the friction–load behaviour is expressed by equation (10) with

$$C = K_5''' D^{(1-\gamma)^2} \quad \ldots\ldots(11\,c) \qquad \text{and} \qquad \beta = \gamma(2-\gamma). \quad \ldots\ldots(12\,c)$$

In the above derivation, the variation of the normal pressure over the area of contact has been ignored. If this variation is taken into account the same equations are obtained but the values of the constants K_5 in equations (11 *b*) and (11 *c*) are altered.

2.4. *Contacts between Smooth Specimens producing Circular Apparent Areas of Contact*

For the particular case of smooth curved specimens the true area of contact is equal to the apparent area of contact and is given by equation (3). If the analysis is repeated it is found that the friction-load behaviour is again expressed by equation (10) with equations (11), (11 *a*) and (11 *c*) modified to

$$C \propto D^{2(1-\alpha_1)} \quad \ldots\ldots(11', 11\,a') \qquad\qquad C \propto D^{1-\gamma_1} \quad \ldots\ldots(11\,c')$$

and equations (12), (12 *a*) and (12 *c*) to

$$\beta = \alpha_1 \quad \ldots\ldots(12', 12\,a') \qquad\qquad \beta = \gamma_1 \quad \ldots\ldots(12\,c')$$

Equations (11′, 11 *a*′), (12′, 12 *a*′) are identical with those obtained by Pascoe and Tabor (1955), the assumptions on which their analysis was based having been listed in § 1.

§ 3. DISCUSSION

3.1. *The Variation of the Apparent Area of Contact with Normal Load*

In § 2 it was shown that the apparent area of contact A_A was related to the load by a power law when a sphere was loaded either against another sphere of the same material or against a hard plane surface, and when cylinders were loaded either to produce a circular or an elliptical area of contact. Partial verification of equation (3) is furnished by the results of Lincoln (1952) and of Howell and Mazur (1953).

Lincoln measured the area of contact between a nylon hemisphere and a glass plate in the load range 0–125 g. The apparent area of contact varied with the load according to the equation $A_A = KL^{0\cdot67}$.

Howell and Mazur measured the area of contact between a glass plate loaded against cylindrical textile fibres which had been wrapped around a hard cylinder of larger radius. This produced elliptical contact areas which varied with the load according to the equation $A_{\mathrm{A}}=KL^{\alpha_1}$. For the materials cellulose acetate, viscose rayon, drawn and undrawn nylon α_1 was found to lie between 0·72 and 0·75.

This verification of equation (3) has been described as partial to the extent that the experimental data quoted above were obtained with fixed values of the specimen diameter D. The most complete investigation, to date, has been undertaken by Pascoe and Tabor (1955) who indented a flat specimen of nylon with spherical-ended indenters of four different diameters. Their results were in close agreement with Meyer's hardness laws, of which equation (3) is a statement. It should be noted that all previous experimental verification of these laws has been obtained with metals (see Tabor 1951).

Thus the experimental data quoted above are evidence in support of the validity of applying assumption (i) and equation (1) to non-metals.

3.2. *The Real and Apparent Areas of Contact*

It follows from assumption (iii) that the true area of contact, A_{T}, is given by F/S_0, hence when $\alpha_1=\alpha_2=1$, i.e. when the specimen becomes plastic in bulk, we find from equations (10), (11) and (12) $A_{\mathrm{T}}=K_5L/S_0=L/p_0=A_{\mathrm{A}}$ since $p_0=L/A_{\mathrm{A}}$.

Thus, when the apparent area of contact is circular and $\alpha_1=\alpha_2=1$ the true and apparent areas of contact are equal, dependent only on the load and the Meyer hardness, and therefore are independent of the size of the contacting bodies. This result was found experimentally by Bowden and Tabor (1939) who measured the electrical resistance between perpendicularly crossed metal cylinders in contact and between a sphere in contact with a flat metal surface. From their data they were able to calculate the true areas of contact. Similarly with flat surfaces in contact, provided the load is sufficient for the condition $\alpha_2=1$ to be valid, the total true area of contact is equal to L/p_0 and is independent of the number of asperities in contact, the size of the asperities and the apparent area of contact. Experimentally it was found that the electrical resistance, and hence the true area of contact between metal flats in contact, was independent of the apparent area of contact.

3.3. *The Dependence of the Coefficient of Friction of Non-metals on the Load*

The considerations given in § 2 are seen to lead to the relation $F=CL^{\beta}$ which has been found empirically by Bowden and Young (1951) for clean diamond sliding on diamond, by Lincoln (1952) for nylon sliding on nylon, by Guthrie and Oliver (1952) for viscose rayon sliding on viscose rayon, by Howell and Mazur (1953) for cellulose acetate, viscose rayon and nylon fibres, each fibre sliding on another of the same material, by Chapman, Pascoe and Tabor (1955) for urea-formaldehyde resin, by Pascoe and Tabor (1955) for nylon sliding on nylon, by Lord (1955) for cotton fibres on cotton and is implied by the results of Lyne (1955) obtained with cellulose acetate yarn sliding round a chromium plated cylinder.

The extent of agreement between theory and experiment may be assessed by using values of α_1, where these are known from static loading experiments, to calculate the corresponding values of β (using equations (12) or (12 *a*)) shown in column (4) of the table and assuming $\alpha_1 = \alpha_2 = \alpha$. These values may then be compared with the experimentally observed values of β shown in column (3). In column (5) the values of β are calculated using equation (12′, 12 *a*′) and the agreement with the experimental values in the last two examples is quite good. It would seem that in these latter cases the assumption that the specimens are rough is not applicable.

(1)	(2)	(3)	(4)	(5)	(6)
Cellulose acetate	0·75	0·96	0·94		a
Viscose rayon	0·73	0·91	0·93		a
Drawn nylon	0·72	0·80	0·92	0·72	a
Undrawn nylon	0·73	0·90	0·93		a
Urea-formaldehyde	0·74	0·90	0·93		b
Nylon	0·67	0·67		0·67	c
Nylon	0·74	0·74		0·74	d

(1) Material; (2) α_1; (3) β; (4) $\beta = \alpha(2-\alpha)$; (5) $\beta = \alpha$; (6) source.

a, Howell and Mazur 1953; b, Chapman, Pascoe and Tabor 1955; c, Lincoln 1952; d, Pascoe and Tabor 1955.

The calculated and experimental values of β are seen to agree to within about 3% in all but the case of drawn nylon, which value is as poorly fitted by equation (12 *a*) as by equation (12 *a*′). This difference between the values of β for drawn and undrawn nylon is all the more surprising when it is remembered that Pascoe and Tabor (1955) used both drawn and undrawn nylon in their experiments and obtained similar values of β with both fibres.

The frictional determinations made by Lincoln (1952) and by Pascoe and Tabor (1955) were both made with two different geometrical arrangements, viz. perpendicularly crossed nylon cylinders and a nylon hemisphere sliding on a flat nylon surface. In both investigations a single friction law was obtained, $F \propto L^{0\cdot67}$ in the former and $F \propto L^{0\cdot74}$ in the latter. Thus, these results confirm, experimentally, the identity of equations (12′) and (12 *a*′).

Lord (1955) has measured the fibre–fibre friction of cotton using fringes of fibres sliding against each other and using a fringe sliding between two pads of parallel fibres. There is greater opportunity for lateral movement of the fibres in the former method than there is in the latter, hence the angle of inclination of the fibres during sliding may be expected to have been smaller in the fringe–pad system. Although the theory has been derived for cylindrical fibres, and cotton fibres, in general, are irregular in cross section, being convoluted along their length, it is, nevertheless, interesting to see the extent to which the theory agrees with experiment in this case. For the fringe–fringe system it might be expected that β will be given by equation (12 *b*), while for the fringe–pads system, β should approximate to the value given by equation (12 *c*). Experimentally, it was found that the first system gave a value $\beta = 0\cdot886$ while the second system gave $\beta = 0\cdot819$. If these values are substituted in equations (12 *b*) and (12 *c*) respectively the value of α obtained in each case is 0·73. Considering the difference between the idealized model used in the derivation of equations (12 *b*) and (12 *c*) and an actual cotton fibre, the agreement between the derived values of α is rather surprising.

3.4. *Amontons' Laws of Friction*

For the special case in which the stress at the asperities becomes constant (i.e. independent of strain) when slip occurs, the value of the index x_2 is zero. Then $\alpha_2[=2/(2+x_2)]=1$, so that in this case (i) both equation (7) and equation (10) become $F=K_3L$, i.e. Amontons' first law of friction is obeyed ($K_3=\mu$) and (ii) in equation (6) the index $1-\alpha_2=0$, i.e. the frictional force becomes independent of the apparent area of contact (Amontons' second law) and independent of the roughness.

Similarly, as shown in § 2.2 (ii), $C(=K_3)$ becomes independent of the roughness and of the specimen diameter (and hence of the apparent area of contact, equation (3)).

Thus, for both flat and curved specimens, both Amontons' laws of friction are obeyed when $\alpha_2=1$ and, in addition, the frictional force then becomes independent of the roughness provided this is not excessive.

It should be noted that for this to occur, it is necessary only that $\alpha_2=1$, *irrespective of the value of* α_1, i.e. although the applied normal load may be insufficient to produce the condition $\alpha_1=1$ Amontons' laws will still be obeyed provided the load is sufficient for the condition $\alpha_2=1$ to be obeyed. This is illustrated by the results obtained by Whitehead (1950) who found that the coefficient of friction of aluminium sliding on aluminium and of silver sliding on silver remained constant when the normal load was varied from the order of 0·02 g to the order of 2000 g. With the experimental arrangement adopted, viz. a hemispherical slider (radius 0·04 cm) in contact with a flat surface, the transition from $\alpha_1=0{\cdot}67$ to $\alpha_1=1$ occurs over the load range from about 0·17 g to about 25 g for aluminium† and from about 2·9 g to about 450 g for silver†, i.e. within the range of normal loads over which the coefficient of friction was found to be independent of the normal load.

Lincoln (1952) has calculated that the loads necessary for the onset of plasticity with nylon are greater than those with steel, at equivalent radii, by a factor of 10^4 and, presumably, the loads necessary for full plasticity will be increased by a factor of similar magnitude. This would account for the observation that with nylon in particular, and non-metals in general, the loads necessary for Amontons' laws to be obeyed are higher than those necessary with metals. Thus, the difference in the typical behaviour of metals and non-metals is determined by the fact that in one case the stress at the asperities is independent of the strain, while in the other case the stress is strain-dependent when slip of the surfaces occurs.

If, as may occur with non-metals when the normal load is sufficiently great, the stress as the asperities becomes independent of strain, then Amontons' laws of friction are obeyed irrespective of the value of α_1. Thus (Spurr, private communication) the coefficient of friction was shown to be independent of the normal load when a clean steel ball was slid against a clean glass plate in the load range 40–2000 g. By repeating the experiments using smoked glass, the track widths at different loads were obtained and these were shown to be proportional to the cube root of the normal load, i.e. the deformation, in bulk, was elastic. These results show that Amontons' laws were obeyed at normal loads which were insufficient to produce the condition $\alpha_1 = 1$, the condition of bulk plasticity. King and Tabor (1954) found that the coefficient of friction was independent of the normal load when steel was slid on rock salt with normal loads ranging from 25 to 1500 g, which indicates that the condition $\alpha_2 = 1$ obtained during these experiments. This was confirmed by their measurements of the track width at the various loads used when it was found that the square of the track width was proportional to the normal load, i.e. that α_1 was equal to unity. Since the condition for $\alpha_1 = 1$ requires a load greater than that necessary for the condition $\alpha_2 = 1$, it follows that over the range of loads used the condition $\alpha_2 = 1$ was valid. Shooter and Tabor (1952) found a similar proportionality between the square of the track width and the normal load when a hard steel hemisphere was slid over the surface of several high polymer materials at normal loads sufficiently high for the coefficient of friction to become independent of the normal load.

---

† These values are calculated for static loading conditions. Because of the possible interaction of the normal and tangential stresses during friction measurements it is possible for the transition from $\alpha_1=0{\cdot}67$ to $\alpha_1=1$ to occur at normal loads lower than these.

From the above discussion it might be expected that at sufficiently low normal loads, when deformation at the asperities is not completely plastic, metals should exhibit frictional characteristics normally associated with non-metals. This has been observed by Martin and Mittelmann (1946) with tungsten wire sliding round a steel cylinder, when it was found that Amontons' law was not obeyed.

It is now apparent that the description of frictional behaviour as 'metallic' and 'non-metallic' is somewhat misleading; the frictional behaviour of a particular solid will be determined primarily by the shape of the stress–strain curve of the material in compression under the particular experimental conditions and by the load range within which the friction observations are made.

3.5. *The Dependence of the Frictional Force on Specimen Roughness*

In §3.4 it has been shown that when $\alpha_2 = 1$ the frictional force becomes independent of the surface roughness. In general $\alpha_2 = 1$ when one metal surface is slid over another, so that, provided the roughness is not excessive, the friction between metals might be expected to be independent of roughness. This has been noted experimentally by Strang and Lewis (1949) and by Bowden and Tabor (1950).

From equation (6) it can be seen that provided $\alpha_2 \neq 1$, the frictional force depends on the roughness factor ω. On the models postulated, the maximum value of h is $\Delta/2$, hence the roughness $\rho < 1$ and therefore $\omega = 1/\rho(1-\rho)$ decreases as the roughness increases, i.e. the frictional force decreases as the roughness of the surface increases, for a fixed normal load and a fixed apparent area of contact.

Experimental evidence of the effect of roughness has been found by King (1950) who showed that the friction between a wool fibre and cylinders of various materials decreased on scratching the surface of the cylinder. Lincoln (1952) found similar evidence when nylon was rubbed against a stainless-steel cylinder, the coefficient of friction decreasing as the roughness of the cylinder increased.

Similarly, Roth, Driscoll and Holt (1942) with dry rubber, and Denny (1953) with lubricated rubber, have shown that when flat rough rubber specimens are slid on smooth tracks, the friction is lower at low loads than when smooth rubber specimens are used. (At high normal loads, Denny has shown that this difference in frictional force disappears, which is consistent with the view that the whole of the apparent area is in contact with the track.)

The theory given above has revealed that while the parameter C, and hence the frictional force, is dependent on the roughness, the index β is independent of the roughness. Because of the 'scale-like' surface structure of animal fibres, the roughness along a fibre in the direction root-to-tip of scale is different from that in the direction tip-to-root, and hence, while C should exhibit a directional dependence, β might be expected to be independent of direction. This has been investigated by Lincoln (1954) with wool fibre, when it was found that although the frictional force was dependent on direction, the value of β was independent of direction.

Thus the adhesion theory, in conjunction with considerations of the influence of multiple contacts on the deformation of asperities, offers an explanation of the dependence of the frictional force on surface roughness, at least for surfaces which approximate to the models postulated in § 2. The Coulomb theory of friction attributes the frictional force to the work necessary to raise one surface over the roughnesses of the other (e.g. Bikerman 1944, 1948), the coefficient of friction being equal to the slope of an asperity in the direction of sliding. Hence a corollary to the Coulomb theory is that the frictional force increases as the roughness of the surfaces increases, whereas, as shown above, according to the adhesion theory, the frictional force is independent of surface roughness when Amontons' laws are obeyed, and decreases with increase in roughness when Amontons' laws are not obeyed. Thus the experimental evidence of the influence of surface roughness on friction is seen to be in accordance with predictions based on the adhesion theory of friction and contrary to the predictions of the Coulomb theory. Only when the surfaces are excessively rough and the physical properties of the rough surface are such that the sliding member does not 'plough' through the asperities will there be a Coulomb component to the measured frictional force (see, for example, the equation derived by Merchant 1940). The adhesion and Coulomb components of the friction are illustrated by some results of Baird and Mieszkis (1955) who investigated the friction of nylon yarns running against alumina surfaces having five different degrees of roughness. With lubricated nylon yarns the coefficient of friction decreased initially with increase in roughness and then increased with further increase in roughness.

3.6. *The Dependence of the Frictional Force on the Apparent Area of Contact*

When the roughness of the surface is uniform equation (7) gives $F=KA_{\mathrm{A}}{}^{1-\alpha_2}L^{\alpha_2}$. If A_{A} is independent of the normal load (this is approximately true for flat specimens) then $\mu=F/L=KA_{\mathrm{A}}{}^{1-\alpha_2}/L^{1-\alpha_2}=K/p_{\mathrm{A}}{}^{1-\alpha_2}$ where p_{A} is the apparent normal pressure.

For a linearly elastic material (e.g. rubber at light loads) $\alpha_2=2/3$, i.e. $\mu=Kp_{\mathrm{A}}{}^{-1/3}$. This equation was found empirically by Schallamach (1952) for flat rubber specimens sliding on plate glass at normal loads ranging from 1 kg to 30 kg.

Howell (1951) varied the angle of inclination of two crossed nylon fibres and found that the frictional force increased as the angle of inclination decreased, i.e. as the apparent area of contact increased.

Gralén (1952) quotes results obtained with Terylene fibres which suggested an increase in the coefficient of friction with increase in fibre diameter. In view of the small variation in fibre diameter (13·5–18·3 μ), Gralén felt that this effect required further experimental verification. This has been furnished by the experiments of Pascoe and Tabor (1955). They measured the friction of crossed nylon fibres of three different diameters and the friction between a nylon hemisphere and a flat nylon surface. The specimen diameter was varied over a range of 320 : 1 and in this range the dependence of frictional force on specimen diameter was in accordance with equations (11′) and (11 *a*′).

Thus, the experimental evidence quoted indicates that for non-metals, the frictional force increases as the apparent area of contact increases.

§ 4. CONCLUSIONS

The theory given above has shown that the surface friction of solids can be explained in terms of the adhesion theory of friction, the differences in behaviour noted at low and at high loads being attributable to differences in the mode of deformation of solids at low and at high loads. This is in accordance with the findings of Shooter and Tabor (1952) who concluded that the most likely explanation of the observed increase in the coefficient of friction at light loads resided 'in the nature of the deformation process'.

An important fact which has emerged during this analysis is that, provided the roughness of the surfaces is not excessive, the frictional force is dependent on the roughness only when conditions are such that Amontons' laws are not obeyed, the frictional force being independent of the roughness when Amontons' laws are obeyed. When the surfaces are excessively rough, or when one surface 'ploughs' through the other (Bowden, Moore and Tabor 1943, Shooter and Tabor 1952) there will be components of the frictional force additional to that due to the adhesion at the points of contact.

Finally, it should be noted that the present treatment has been simplified by making assumption (v), viz. that the normal and tangential stresses are independent of each other. The implication of this assumption is that the area of contact produced by the normal load remains unaltered as the tangential force is applied, and deductions from theories incorporating this assumption are in satisfactory qualitative agreement with experiment when applied to most materials, the quantitative agreement being poorer. When attempts are made to explain the frictional behaviour of very soft (low melting point) metals at room temperature it is found (McFarlane and Tabor 1950) that this assumption must be abandoned and allowance must be made for the interaction of the applied stresses. Hence it may be concluded that to explain the frictional behaviour, at elevated temperatures, of materials which show marked softening at these temperatures (e.g. thermoplastic polymers, metals) the interaction of the normal and tangential stresses will have to be taken into account by means of the appropriate yield criterion for failure under the application of combined stresses.

ACKNOWLEDGMENTS

The author wishes to thank members of the staff of the British Cotton Industry Research Association and, in particular, Mr. C. Mack for helpful comments on and criticism of the manuscript.

REFERENCES

ARCHARD, J. F., 1953, *Nature, Lond.*, **172,** 918.

BAIRD, M. F., and MIESZKIS, K. W., 1955, *J. Text. Inst.*, **46,** P101.

BIKERMAN, J. J., 1944, *Rev. Mod. Phys.*, **16,** 53; 1948, *Lubric. Engng*, **4,** 208.

BOWDEN, F. P., MOORE, A. J. W., and TABOR, D., 1943, *J. Appl. Phys.*, **14,** 80.

BOWDEN, F. P., and TABOR, D., 1939, *Proc. Roy. Soc.* A, **169,** 391; 1950, *The Friction and Lubrication of Solids* (Oxford : University Press) pp. 78, 175.

Bowden, F. P., and Young, J. E., 1951, *Proc. Roy. Soc.* A, **208**, 444.
Chapman, J. A., Pascoe, M. W., and Tabor, D., 1955, *J. Text. Inst.*, **46**, P3.
Denny, D. F., 1953, *Proc. Phys. Soc.* B, **66**, 721.
Gralén, N., 1952, *Proc. Roy. Soc.* A, **212**, 491.
Gralén, N., Oloffson, B., and Lindberg, J., 1953, *Text. Res. (J.)*, **23**, 623
Guthrie, J. C., and Oliver, P. H., 1952, *J. Text. Inst.*, **43**, T579.
Hertz, H., 1881, *J. reine angew. Math.*, **92**, 156.
Holm, R., 1938, *Wiss. Veröff. Siemens*, **17**, 38.
Howell, H. G., 1951, *J. Text. Inst.*, **42**, T521.
Howell, H. G., and Mazur, J., 1953, *J. Text. Inst.*, **44**, T59.
King, G., 1950, *J. Text. Inst.*, **41**, T135.
King, R. F., and Tabor, D., 1954, *Proc. Roy. Soc.* A, **223**, 225.
Lincoln, B., 1952, *Brit. J. Appl. Phys.*, **3**, 260; 1953, *Nature, Lond.*, **172**, 169; 1954 *J. Text. Inst.*, **45**, T92.
Lodge, A. S., and Howell, H. G., 1954, *Proc. Phys. Soc.* B, **67**, 89.
Lord, E., 1955, *J. Text. Inst.*, **46**, P41.
Lyne, D. G., 1955, *J. Text. Inst.*, **46**, P112.
McFarlane, J. S., and Tabor, D., 1950, *Proc. Roy. Soc.* A, **202**, 244.
Martin, A. J. P., and Mittelmann, R., 1946, *J. Text. Inst.*, **37**, T269.
Merchant, M. E., 1940, *J. Appl. Phys.*, **11**, 230.
Nadai, A., 1931, *Plasticity* (New York : McGraw-Hill), p. 83.
Pascoe, M. W., and Tabor, D., 1955, *Research, Lond.*, **8**, S15.
Roth, F. L., Driscoll, R. L., and Holt, W. L., 1942, *J. Res. Nat. Bur. Stand.*, **28**, 439.
Schallamach, A., 1952, *Proc. Phys. Soc.* B, **65**, 657.
Shooter, K. V., and Tabor, D., 1952, *Proc. Phys. Soc.* B, **65**, 661.
Strang, C. D., and Lewis, C. R., 1949, *J. Appl. Phys.*, **20**, 1164.
Tabor, D., 1948, *Proc. Roy. Soc.* A, **192**, 247; 1951, *The Hardness of Metals* (Oxford University Press) pp. 8–14.
Whitehead, J. R., 1950, *Proc. Roy. Soc.* A, **201**, 109.

Halide Activated Halide Phosphors

By J. W. ELDER[†]

Physics Department, Canterbury College, Christchurch, New Zealand

MS. received 21*st February* 1956, *and in revised form* 9*th April* 1956

Abstract. New absorption peaks have been observed in the long wavelength tail of the fundamental absorption band of pure alkali halides, when a small amount of another alkali halide is added as an impurity. Results are given which suggest that the electronic levels involved in the optical process are associated with the activator ion and that the mechanism may be described by a configurational coordinate model.

§ 1. INTRODUCTION

THE understanding of absorption and emission in complex phosphors may be found by studies of simple phosphors. Such a phosphor would contain localized impurity centres. Thallium-activated halides have already been investigated. Alkali halides, containing a small proportion of halide impurity in substitutional positions, have been found to have new absorption and emission bands below the fundamental limit of the pure crystal. These are among the closest approaches to an ideal phosphor.

§ 2. EXPERIMENTAL METHOD

Crystals were grown, by the Kyropoulos technique, from melts of salts MX containing 1 mole per cent of a salt MY, where $M = Na^+, K^+, Rb^+$ and $X, Y = F^-, Cl^-, Br^-, I^-$. Some of the samples were grown directly on to a 2 cm length of 4 mm diameter nickel rod, screwed into the end of the cooling mechanism. The tip of the rod was conical with a groove just above the cone. Boules were multicrystalline, but usable pieces of crystal were easily obtained.

Absorption measurements were made with a Spekker photometer, using a Hilger quartz ultra-violet monochromator, an IP28 photomultiplier as detector, and, because of its nearly constant intensity across the spectrum, an HF6 hydrogen discharge lamp as a source. The ground base of the prism, in the above monochromator, produces stray light by scattering of light internally reflected within the prism. The form of the scattered spectrum is the same as that of the incident spectrum but is 10^{-3} times as intense. Mercury arc sources have, over the range 2100–6000 Å, an intensity range of about 10^3 and thus below 2700 Å produce errors of the order of 100%. Entrance and exit slits were set at 0·25 mm but the reported half-widths have been corrected (Brodersen 1954).

A phase sensitive coherent null detector, similar in principle to that of Liston and White (1950), was used. The two beams of the Spekker photometer were chopped at 40 c/s. The square wave output of the photomultiplier, amplified by

---

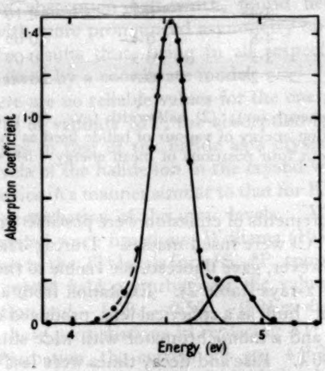
[†] Now at the Cavendish Laboratory, Cambridge.

5×10^4, was fed into a ring modulator, phased by an electronic switch (*Waveforms* 1949). The switch was triggered by symmetrically spaced pulses produced in a magnetic pickup coil when two steel nails attached to the aluminium chopper disc passed over it. The output from the ring modulator was integrated in an RC circuit to give a band width of 0·03 c/s. The measurements of optical density were reproducible to better than 0·002 over the range of densities of 0·2 to 2·0 at wavelengths down to 2150 Å. Relative intensities were measured with the same detector, by generating at the photomultiplier anode a known square wave, 180° out of phase with the chopped light beam.

The background absorption, due to the solvent, was eliminated by subtracting the absorption of the pure salt, MX, from that of the activated salt, MX + Y. This is complicated by two independent effects. The absorption curves of different samples of a given salt show different levels of a wavelength independent absorption, and secondly variations in the shape of the absorption rising at short wavelengths. Absorption outside the region of the impurity peaks—above 5·5 ev and below 3 ev—has been shown to be entirely due to the solvent.

§ 3. RESULTS

The impurity peaks are well fitted by gaussian curves but are slightly asymmetric in the tails. Complete spectra were analysed graphically into gaussian components by mounting a lantern slide of a set of gaussian curves in a projector so that the slide could be rotated about the two axes perpendicular to that of the projector. The absorption energies E and the half-widths are given in electron volts in table 1, together with the values of the peak absorption coefficients and for comparison the absorption energy of MY as a vapour (Barrow and Caunt 1953, Müller 1927). The exact impurity content in the crystals has not been determined but it is probably near $1:10^3$. An asymmetry was apparent only in the 4 to 4·5 ev tails of the strong peaks of NaCl: F, Br, I; NaBr: Cl; KCl: Br; RbCl: Br. For these peaks a good fit is obtained with an expression of the form



Absorption Spectrum of KCl : Br.

$$P = (\exp - bx^2)/(1 - ax)^{1/2}, \quad x = E - E_{\max}$$

(Williams 1951), where $a=2{\cdot}5$, 4·5, 2·0, 2·5, 4·5, 3·0 ev^{-1}, for the above crystals respectively. This accounts for all but a very small amount of absorption, occurring to 0·1 ev below the tail of the above curve. In the figure the experimentally determined absorption coefficient (=ln transmission/cm), is shown plotted against energy for a typical curve, that of Br$^-$ in KCl. The gaussian curves best fitting the data are also shown.

The variation of the half-width W of the 4·5 ev peak of KCl : Br was determined from $T=100°$K to 700°K. Below 200°K a departure from Mott's expression $W=cT^{1/2}$ was found. The experimental data were fitted by the expression $W=C(\theta \coth \theta/T)^{1/2}$ (Williams and Hebb 1951) by least squares, yielding $\theta=(90 \pm 20$ s.d.)°K, corresponding to a zero point energy $k\theta=0{\cdot}008$ ev. The energy of the absorption peak was almost independent of temperature.

Table 1. Absorption Characteristics of Halide Activated Halides at 300°K

Crystal	(1)	(2)	(3)	(4)
NaCl : F	4·90	0·56	2·76	5·6
NaBr : Cl	4·8	0·6	2·07	5·1
NaCl : Br	4·90	0·38	2·92	4·85
	4·49	0·32	2·30	4·45
KCl : Br	4·83	0·22	0·23	4·88
	4·53	0·21	1·47	4·49
RbCl : Br	4·83	0·36	1·31	4·82
	4·53	0·37	2·46	4·43
NaCl : I	4·85	0·55	1·98	4·80
	3·60	0·25	0·18	3·80
	3·45	0·15	0·16	—
KBr : I	4·8	0·4	0·09	—
	3·5	0·5	0·09	3·8
	3·25	0·3	0·09	—

(1) Peak absorption energy (ev); (2) half-width (ev); (3) peak absorption coefficient (cm^{-1}); (4) peak absorption energy of vapour of halide used as the impurity (ev).

The peaks cluster about four positions of mean energy 4·84, 4·52, 3·55, 3·35 electron volts.

Quantitative measurements of emission were possible with three samples only, of which KBr : Cl, KI : Cl were fused masses. During irradiation by ultra-violet light most samples, however, gave fluorescence visible to the eye, as did the above three samples to 45 kv x-rays (table 2). Excitation from a mercury arc, using a water filled 'black-light' bulb as a spherical lens, produced excitation in the range between 3000–4000 Å, and a monochromator with wide slits produced excitation in the range 2000–3000 Å. Rise and decay times were less than 50 milliseconds. A NaCl : 10% NaI saturated solution has a blue fluorescence, but changes to yellow when evaporated and strongly heated. Thus the luminescent behaviour of these phosphors is similar to that of KCl : Tl, in solution or deposited as micro-crystals (Pringsheim 1942).

Experiments with a field of 2000 volt cm^{-1} applied to samples 1 cm × 1 cm × 0·1 cm, excited by the full spectrum of a medium pressure mercury arc, indicated that possible photo-currents were less than 10^{-14} amp.

Table 2. Emission Characteristics of Halide Activated Halides at 300°K under 3000–4000Å excitation

Sample	Position (ev)	Width (ev)	Strength
RbCl : Br	2·38	0·46	Strong
KBr : Cl	2·45	0·39	Strong
KI : Cl	2·30	0·38	Strong
NaCl : F, NaCl : Br, KCl : Br, KBr : I	2·5	—	Very weak
NaCl : I	2·0	—	Weak

Under 2000–3000 Å excitation all samples gave weak emission near 2·5 ev.

§ 4. Discussion of Results

All crystalline materials are capable of electronic conduction if free electrons can be produced in them. That the above crystals do not conduct implies that the electronic levels are not in or close to a conduction band. Localized energy levels are further implied by the independence of peak transition energy on both temperature and the solvent. Of the two basic models used to describe crystal phosphors, the band and the configurational coordinate model, the former is thus not applicable to the impurity levels so far discovered.

For a trapped electron the energy of transition between two states usually depends strongly on a small number of nearest neighbour coordinates. Williams (1951, 1953) has successfully applied this coordinate model to KCl : Tl obtaining the expressions for *P*, *W* given above. He also finds that the energy levels are quadratic in the nearest neighbour distance, the curvature of excited states being smaller near their minima than that of the ground state. This implies firstly greater emission than absorption half-width, found here for RbCl : Br, and secondly line shape with more pronounced asymmetry on the small energy side, as found above. The results thus, being in all respects similar to those for KCl : Tl, can be explained by a coordinate model.

Unfortunately there are no reliable values for the energy levels of halide ions other than in the form of vapours MY, the absorption spectra of which bear a remarkable resemblance to that of the halide as a crystal impurity MX + MY. Assuming that the levels of the halide ion in the crystal will be perturbed by the surrounding crystal lattice in a manner similar to that for KCl : Tl a rough estimate can be made of the perturbation of the ionic levels. Allowing for the angular dependence of the excited wave functions, Williams (1951, 1953) finds that the principal perturbations to the Tl^+ levels for a $^1S_0$–$^3P_1$ transition are the repulsion, van der Waals, and crystal field perturbations, the values being $-0{\cdot}44$, $-0{\cdot}31$, $-0{\cdot}91$ ev for absorption and $-1{\cdot}10$, $-0{\cdot}50$, $-0{\cdot}91$ ev for emission. Provided the major change in these perturbations in replacing Tl^+ by Y^- is to change the sign of the field perturbation, the corrections are $+0{\cdot}2$ ev for absorption and $-0{\cdot}7$ ev for emission, that is, the change in the ionic levels is small for absorption and large for emission, and we would expect similar results for the absorption for the impurity either in the crystal or as a vapour. Similar results are found for solutions of MY in water. It is well known that the ground state of a halogen atom is a doublet the separations being 0·03, 0·11, 0·44, 0·94 ev for F, Cl, Br, I. In halides containing Br, I impurities there is a multiple structure, the separations being 0·3 – 0·4, 1·3 – 1·4 ev. Thus the energy levels are those of the activator ions.

§ 5. CONCLUSION

On absorption of a photon the crystal becomes excited but after a very short time the excitation becomes localized at an impurity centre, an impurity electron being raised to a discrete excited perturbed level of the impurity ion. Following a rearrangement of the lattice to a position of minimum energy the impurity electron falls to the ground state by emission of a photon. These transitions may be described by a coordinate model with a classical distribution of atomic configurations, characteristic of the ground state except below 200°K when the zero point energy must be considered. The results show that the halide activated halide phosphors have properties similar to thallium activated alkali halides and that the coordinate model applies to a greater range of phosphors than has hitherto been supposed.

ACKNOWLEDGMENTS

The writer wishes to thank Mr. D. Shipley for suggesting the use of the magnetic pickup, Dr. N. K. Pope for his help and encouragement and the Shirtcliffe Fellowship Committee for the award of their Fellowship in 1955.

REFERENCES

BARROW, R. F., and CAUNT, A. D., 1953, *Proc. Roy. Soc.* A, **219**, 120.
BRODERSEN, S., 1954, *J. Opt. Soc. Amer.*, **44**, 22.
LISTON, M. D., and WHITE, J. U., 1950, *J. Opt. Soc. Amer.*, **40**, 36.
MÜLLER, L. A., 1927, *Ann. Phys., Lpz.*, **82**, 39.
PRINGSHEIM, P., 1942, *Rev. Mod. Phys.*, **14**, 132.
Waveforms, 1949, Radiation Lab. Series, Vol. 19 (New York : McGraw Hill), pp. 210, 604.
WILLIAMS, F. E., 1951, *J. Chem. Phys.*, **19**, 464; 1953, *J. Phys. Chem.*, **57**, 780.
WILLIAMS, F. E., and HEBB, M. H., 1951, *Phys. Rev.*, **84**, 1181.

On the Theory of Plasma Waves

By F. BERZ†

Department of Electrical Engineering, Imperial College, London

Communicated by D. Gabor; MS. received 2nd June 1955, *and in revised form 9th March* 1956

Abstract. This theory applies to plasmas with arbitrary stationary electron velocity distributions, neglecting the collisions of electrons with charged or neutral particles. The plasma waves can be divided into two groups.

In the first group are the waves which depend on the conditions of excitation. These waves are damped and can have, under suitable initial conditions, any frequency, wavelength and damping, within certain limits. In the second group are the 'proper' plasma waves, i.e. the waves whose dispersion characteristics are not influenced by the initial conditions, but are solely determined by the stationary electron density and the stationary electron velocity distribution f_0 of the plasma. They include the whole class, if any, of amplified plasma waves, and also some definite types of damped and undamped waves. The amplified plasma waves can exist only for special types of electron velocity distributions. 'Proper' damped or undamped plasma waves always arise, whatever the initial conditions at $t=0$. The oscillating velocity distribution for these waves has two components, one which varies in space and time as the electric field, and the other which is carried along with the electrons.

For electron distributions f_0 of the Maxwellian type, all the plasma waves are damped, except for $\omega \simeq \omega_0$, where $\omega_0=(4\pi \boldsymbol{n}\boldsymbol{e}^2/\boldsymbol{m})^{1/2}$. This damping is due to the intrinsic properties of the plasma. For $\omega_0<\omega<1{\cdot}05\omega_0$ the damping of the 'proper' plasma waves is small, and the corresponding wavelengths cover a wide range $30\lambda_D<\lambda<\infty$ (λ_D = Debye length). The damping becomes large for $\omega>1{\cdot}15\omega_0$, $\lambda<20\lambda_D$. Langmuir's simplified theory gives a good first approximation ($\omega=\omega_0$, λ arbitrary) to these results.

In the well-known theory of Bohm and Gross no initial conditions are taken into account, and it is assumed that the oscillating density distribution has no component $\bar{f}^*$ carried along with the electrons. Their dispersion equation is identical with the general dispersion relation for amplified waves, but for non-amplified or damped waves it would correspond to special conditions of excitation. Contrary to the conclusions of Bohm and Gross this equation has no solutions, except for waves which propagate faster than any electron. In all other cases plasma waves can exist only if there is a component $\bar{f}^*$.

General Introduction

Plasma-electron waves are longitudinal electric oscillations of high frequency, which can arise in ionized media. These waves are accompanied by a periodic motion of electrons, while at these high frequencies the ions remain practically motionless, due to their much larger mass.

---

† Now at Mullards Research Laboratories, Salfords, Redhill, Surrey.

Plasma waves exhibit a very wide variety, depending on the static field configurations, the boundary conditions and the velocity distributions of the electrons in the ionized medium.

We shall consider in this paper only one-dimensional problems, in regions without static fields.

The velocity distributions of electrons in the absence of oscillations is taken to be continuous. Particular attention is given in some parts to the case when this distribution is of the Maxwellian type, such as can be found for instance in the positive column of a gas discharge.† No relativity effect has been taken into account, since below 10^8 deg K the average velocity of electrons in thermal equilibrium is small compared with the velocity of light $\mathbf{c}$.‡

The damping effect of collisions of the electrons with other electrons, ions or neutral particles is neglected in the present paper. This is justified for the usual values of charge densities and for pressures such that the collision frequency between any given electron and the neutral particles is small compared with the plasma frequency, which commonly occurs for low pressure discharges (see Berz 1955, p. 94, Bohm and Gross 1949 b).

The main theoretical investigations by previous authors are summarized in the table.

	Assumptions	Results
I	All the fluctuating quantities vary as $e^{i(kx-\omega t)}$. The electrons are motionless in the absence of oscillations.	One single frequency of oscillation $\omega=\omega_0\simeq 9000\sqrt{n}$ (n=electron density cm^{-3}), with any wavelength.
II	All the fluctuating quantities vary as $e^{i(kx-\omega t)}$. The electrons have a thermal velocity distribution.	Continuous frequency spectrum for $\omega\geqslant\omega_0$. The waves are undamped. $\omega^2\simeq\omega_0^2+3k^2\overline{u^2}$. ($(\overline{u^2})^{1/2}$=r.m.s. electron velocity in the x direction).
III	Initial conditions at $t=0$ are taken into account. The electrons have a thermal velocity distribution.	Continuous frequency spectrum, for $\omega\geqslant\omega_0$. The waves are *damped*, i.e. k is complex. $\omega^2\simeq\omega_0^2+3(\mathscr{R}k)^2\overline{u^2}$.

I, Langmuir and Tonks 1929; II, Vlassov 1945, Bohm and Gross 1949 a, b; III, Landau 1946.

It is seen from this table that there are considerable differences between the results of the various authors. The aim of the present paper is to give an elucidation of these discrepancies, and to provide a better theoretical understanding of the whole question.

---

† The concept of the velocity distribution function f is based on the assumption that the discrete electrons of charge $\boldsymbol{e}$ and mass $\boldsymbol{m}$ can be replaced by a continuous, electrically charged fluid, characterized in the phase space by a charge density $\boldsymbol{e}f$, and a specific mass $\boldsymbol{m}/\boldsymbol{e}$. This model has been commonly accepted and is also used in the present paper. Its validity would however deserve a most careful investigation. The first attempt at this can be found in the interesting pioneering papers of Ecker (1955 a, b, c).

‡ Moreover a simple calculation shows that at 10^6 deg K there would be only one electron with a velocity larger than $\boldsymbol{c}/2$, per e^{336} electrons present. This last figure exceeds the total number of electrons of the Universe, as given by Eddington !

§ 1. ON THE ABSENCE OF A DEFINITE DISPERSION EQUATION FOR PLASMA WAVES

1.1. *Introduction*

Let us consider a plasma wave with an electric field which varies as $e^{i(kx-\omega t)}$. It has been pointed out by Landau (1946) and Twiss (1952), on rather general mathematical grounds, that the dispersion relation between k and ω must be strongly affected by the initial conditions. Such conditions are not taken into account in the 'substitution' analysis used by Vlassov and by Bohm and Gross, where it is assumed *a priori* that the oscillating velocity distribution of the electrons also varies as $e^{i(kx-\omega t)}$. This leads to particular solutions, and not to the complete dispersion relation.

The result is somewhat unexpected, since it is well known that the substitution analysis gives the most general solutions in the case of media free of space charge.

It is shown in this section by means of a simple method how the plasma waves are affected by the initial conditions and what is the extent of this influence.

1.2. *Fundamental Equations and Assumptions*

The fundamental equations of longitudinal plasma waves are derived from Boltzmann's transfer equation (which contains the equations of motion and the continuity equation) and Poisson's equation

$$\frac{\partial}{\partial t}(f_0+\tilde{f})+u\frac{\partial}{\partial x}(f_0+\tilde{f})-\frac{e}{m}\tilde{E}\frac{\partial}{\partial u}(f_0+\tilde{f})=\frac{\partial}{\partial t}(f_0+\tilde{f})\bigg|_c \quad \ldots\ldots(1)$$

$$\frac{\partial \tilde{E}}{\partial x}=-4\pi e n\int_{-\infty}^{\infty}\tilde{f}(u,x,t)\,du. \quad \ldots\ldots(2)$$

Gaussian units are used throughout. t, x, u are the time, abscissa and electron velocity in the x direction, n is the equilibrium electron concentration, $\tilde{E}$ the fluctuating electric field intensity, $nf_0(u)$ the electron velocity distribution in the stationary state, i.e. in the absence of the fluctuating field (the integral from $-\infty$ to ∞ of $f_0(u)$ is unity); $n\tilde{f}$ is the fluctuating velocity distribution and $n(\partial/\partial t)(f_0+\tilde{f})\,|_c$ is the variation in the velocity distribution due to collisions.

It is assumed in (1) that no electric field exists in stationary conditions, and in (2) that the stationary electron density is exactly balanced by an equal positive ion density. Further f_0 and $\tilde{f}$ are assumed to be continuous, as well as their derivatives, which appear in (1).

Equation (1) is satisfied for stationary conditions by

$$\frac{\partial f_0}{\partial t}+u\frac{\partial f_0}{\partial u}=\frac{\partial f_0}{\partial t}\bigg|_c.$$

Thus, (1) becomes

$$\frac{\partial \tilde{f}}{\partial t}+u\frac{\partial \tilde{f}}{\partial x}-\frac{e}{m}\tilde{E}\left(\frac{df_0}{du}+\frac{\partial \tilde{f}}{\partial u}\right)=\frac{\partial \tilde{f}}{\partial t}\bigg|_c.$$

We shall further simplify this equation by neglecting the collision term $(\partial\tilde{f}/\partial t)\,|_c$ and the product of the fluctuating variables $\tilde{E}\,\partial\tilde{f}/\partial u$. The first step is justified when the collision frequency is small compared with the plasma frequency, and the second step is justified when the oscillating component $\tilde{f}$ is small compared with f_0.

Thus the fundamental equations (1) and (2) become

$$\frac{\partial \tilde{f}}{\partial t}+u\frac{\partial \tilde{f}}{\partial x}-\frac{e}{m}\tilde{E}\frac{df_0}{du}=0 \qquad \ldots\ldots(3)$$

$$\frac{\partial \tilde{E}}{\partial x}=-4\pi en\int_{-\infty}^{\infty}\tilde{f}(u,\,x,\,t)\,du. \qquad \ldots\ldots(4)$$

In what follows we shall assume that $\tilde{E}$ varies as $e^{i(kx-\omega t)}$. This only means that the field conforms to the usual wave pattern, a necessary assumption if we speak of plasma waves. It does not, however, follow that $\tilde{f}$ should vary also as $e^{i(kx-\omega t)}$, as is made clear below.

The case when k is real will be considered in detail. The case when ω is real could be developed along similar lines.

1.3. *Case when the Wave Number k is Real*

When k is real we can put $\tilde{E}=Ee^{i(kx-\omega t)}$, $\tilde{f}=f(u,\,t)\,e^{ikx}$.

Boltzmann's transfer equation (3) and Poisson's equation (4) become

$$\frac{\partial f}{\partial t}+ikuf-\frac{e}{m}Ee^{-i\omega t}\frac{df_0}{du}=0 \qquad \ldots\ldots(5)$$

$$ik\,Ee^{-i\omega t}=-4\pi en\int_{-\infty}^{\infty}f(u,\,t)\,du. \qquad \ldots\ldots(6)$$

Now (5) is a linear equation of the first order with constant coefficients, whose solution, as obtained by standard methods, is

$$f(u,\,t)=-i\frac{e}{m}E\frac{f_0'(u)e^{-i\omega t}}{ku-\omega}-i\frac{e}{m}C(u)e^{-ikut} \qquad \ldots\ldots(7)$$

where $C(u)$ is a function of u which can have any arbitrary form as far as the mathematical solution of (5) is concerned. In practice $C(u)$ is determined by the initial conditions at $t=0$, as can be seen from (7), and as is further shown in § 2.

Substituting back into (6) we obtain after introducing the plasma angular frequency $\omega_0=(4\pi ne^2/m)^{1/2}$ and multiplying both sides of (6) by $ke^{i\omega t}/\omega_0{}^2iE$

$$\left(\frac{k}{\omega_0}\right)^2=\int_{-\infty}^{\infty}\frac{f_0'(u)}{u-\omega/k}\,du+\frac{k}{E}e^{i\omega t}\int_{-\infty}^{\infty}C(u)e^{-ikut}\,du. \qquad \ldots\ldots(8)$$

This formula in order to be self-consistent should be independent of t.

Let us take in equation (8)

$$C(u)=\frac{EC_0}{\omega/k-u} \qquad \ldots\ldots(9)$$

where C_0 is a constant, and $\mathscr{I}\omega<0$ (damped waves). Applying Jordan's lemma (Whittaker and Watson 1927) and the theorem of residues to the contour formed by the real axis and the semicircle σ, $|u|=R$, in the lower half-plane, we obtain for $R\rightarrow\infty$

$$\int_{-\infty}^{\infty}\frac{e^{-ikut}}{\omega/k-u}\,du=2\pi i\,e^{-i\omega t}.$$

For damped waves (8) becomes

$$\left(\frac{k}{\omega_0}\right)^2=\int_{-\infty}^{\infty}\frac{f_0'(u)}{u-\omega/k}\,du+2\pi ikC_0\dagger \qquad \ldots\ldots(10)$$

---

† A dispersion equation independent of t could also be obtained for any choice of $C(u)$ other than (9), provided $|C(u)|\rightarrow 0$ along σ, and that the only singularity of $C(u)$ in the lower half plane is a pole of order one at $u=\omega/k$.

It is thus seen that, for any given real value of k, any value of ω, such that $\mathscr{I}\omega<0$, satisfies (10), if C_0, which is determined by the initial conditions, is chosen appropriately. In practice the electrons have finite velocities $|u|<u_{\mathrm{M}}<c$ (see footnote ‡ p. 940). Therefore $C(u)$ can take a form similar to (9) only for $|u|<u_{\mathrm{M}}$, and then tapers to zero. It could be shown that this does not alter appreciably the results, provided $|\omega/k|<u_{\mathrm{M}}$ and $\dfrac{\mathscr{I}\omega/k}{u_{\mathrm{M}}-|\omega/k|}\ll 1$. Within these limits waves can propagate with any wavelength, frequency and damping, provided some electrons are present at the wave velocity and a proper choice is made of the initial velocity distribution. Now according to (7)

$$\tilde{f}(u,x,t)=-i\frac{e}{m}E\frac{f_0'}{ku-\omega}\mathrm{e}^{i(kx-\omega t)}+i\frac{e}{m}E\frac{C_0k}{ku-\omega}\mathrm{e}^{i(kx-kut)}. \quad \ldots\ldots(11)$$

The last term represents disturbances carried along with the electron velocity.

If we assume that $\tilde{f}$ varies like $\widetilde{E}$ in space and time, $C_0=0$. This is the case which has been considered by Vlassov (1945) and Bohm and Gross (1949 a, b) and discussed here in § 3.

Consider now amplified waves ($\mathscr{I}\omega>0$). If we choose again $C(u)$ according to (9), it can be easily seen by Jordan's lemma that:

$$\int_{-\infty}^{\infty}C(u)\mathrm{e}^{-ikut}\,du=C_0E\int_{-\infty}^{\infty}\frac{\mathrm{e}^{-ikut}}{\omega/k-u}\,du=0$$

and therefore the dispersion equation becomes

$$\left(\frac{k}{\omega_0}\right)^2=\int_{-\infty}^{\infty}\frac{f_0'(u)}{u-\omega/k}\,du.\dagger \quad \ldots\ldots(12)$$

That this is the most general equation for amplified waves will receive confirmation in § 2. We are back again to the equation obtained with $C_0=0$. Thus only the component of $\tilde{f}$ which varies as the electric field comes into the dispersion equation for amplified waves. This equation is uniquely determined by the stationary characteristics of the plasma and is independent of the initial conditions. It is shown in §§ 2 and 3 that equation (12) has no solution, and therefore no amplified waves can arise, except for special types of velocity distributions f_0.

Similar results could also be derived for the case when ω is real.

§ 2. PLASMA WAVES FOR GIVEN INITIAL CONDITIONS AT $t=0$

2.1. *General Results*

It has been shown in § 1 that the dispersion equation for plasma oscillations depends on the assumptions concerning $\tilde{f}$. It is therefore essential to take the initial conditions into account.

The formal derivation has been done by Landau (1946), Twiss (1951, 1952) and Berz (1955, p. 39) and the results obtained are first briefly reviewed here.

The initial velocity distribution can be put in the form

$$f_0(u)+\int_{-\infty}^{\infty}g_k(u)\mathrm{e}^{ikx}\,dk.$$

---

† This dispersion equation would also be obtained when $C(u)$ is chosen according to the more general conditions given in the footnote of p. 942.

Since the problem has been made linear, the solution will be the integral of the solutions corresponding to each of the k-components. We therefore take an initial velocity distribution at $t=0$, of the form

$$f_0(u)+g(u)e^{ikx} \qquad \text{......(13)}$$

where k is real (>0).

The subsequent behaviour of the plasma can be derived by means of (3) and (4) with the help of Laplace's transforms (see Appendices I and II). It is then seen that the electric field $\widetilde{E}$ has the form:

$$\widetilde{E}=\sum_j A_j \exp\{i(kx-\Omega_j t)\}+\sum_p B_p \exp\{i(kx-\omega_p t)\}. \qquad \text{......(14)}$$

A_j and B_p are constants.†

The first sum corresponds to damped waves which depend on the initial disturbance $g(u)$, since

$$\Omega_j = k u_j \qquad \text{......(15)}$$

where the points u_j are the singularities, if any, of $g(u)$ in the lower half-plane $\mathscr{I}\omega<0$.

The second sum in (14) corresponds to 'proper' plasma waves, whose frequencies ω_p are determined by the static characteristics of the plasma, and are given by the roots of the following equations:

For $\mathscr{I}\omega>0$ (amplified waves)

$$\left(\frac{k}{\omega_0}\right)^2=\int_{-\infty}^{\infty}\frac{f_0'(u)}{u-\omega/k}\,du. \qquad \text{......(16}a\text{)}$$

For $\mathscr{I}\omega=0$

$$\left(\frac{k}{\omega_0}\right)^2=\mathrm{P.}\int_{-\infty}^{\infty}\frac{f_0'(u)}{u-\omega/k}\,du+i\pi f_0'\left(\frac{\omega}{k}\right). \qquad \text{......(16}b\text{)}$$

For $\mathscr{I}\omega<0$ (damped waves)

$$\left(\frac{k}{\omega_0}\right)^2=\int_{-\infty}^{\infty}\frac{f_0'(u)}{u-\omega/k}\,du+2i\pi f_0'\left(\frac{\omega}{k}\right). \qquad \text{......(16}c\text{)}$$

The terms on the right-hand sides of equations (16*b*) and (16*c*) represent the analytical continuation of the function of ω, defined by the integral in (16*a*) for the upper half-plane (Appendix II). The notation P. in (16*b*) stands for the principal value of the integral, according to Cauchy's definition. It can be proved rigorously (Berz 1955, p. 66) that for ω/k large with respect to the average values of u

$$\int_{-\infty}^{\infty}\frac{f_0'(u)}{u-\omega/k}\,du \simeq -\frac{k}{\omega}\int_{-\infty}^{\infty}\left[1+\frac{k}{\omega}u+\left(\frac{k}{\omega}\right)^2 u^2+\dots\right]f_0'(u)\,du$$

$$=\left(\frac{k}{\omega}\right)^2\left[1+2\frac{k}{\omega}\bar{u}+3\left(\frac{k}{\omega}\right)^2\overline{u^2}(1+\epsilon)\right]$$

---

† It is assumed (i) that $f_0(u)$ and $g(u)$ are analytical with poles as only singularities, (ii) that there is a family of curves in the lower half plane, whose distance from the origin tends to infinity, and along which

$$\left|\frac{g(u)}{1-2i\pi(\omega_0/k)^2 f_0'(u)}\right| \to 0 \text{ for } |u| \to \infty$$

(see Berz 1955, p. 39).

where the integral is replaced by its principal value when ω/k is real, and where

$$\bar{u}=\int_{-\infty}^{\infty} u f_0(u)\,du, \qquad \overline{u^2}=\int_{-\infty}^{\infty} u^2 f_0(u)\,du$$

and $|\epsilon|\to 0$ when $|\omega/k|\to\infty$.

Therefore it is seen from (16) that the following asymptotic expansion is obtained for ω/k large

$$\left(\frac{\omega}{\omega_0}\right)^2 \simeq 1+2\frac{k}{\omega}\,\bar{u}+3\left(\frac{k}{\omega}\right)^2\overline{u^2}. \qquad \ldots\ldots(17)$$

It follows from a comparison of (16 *a*) with (8), taking (7) into account, that, for amplified waves, only the component of the fluctuating density distribution $\tilde{f}$ which varies in space and time as the oscillating electric field gives a contribution to the dispersion equation. This agrees with the conclusions of § 1.

For damped waves, however, the component of $\tilde{f}$ which is carried along with the electrons plays an important part. It accounts for the second term on the right-hand side of (16 *c*).

2.2. *Existence and Number of 'Proper' Plasma Waves*

Formulae (16) give the dispersion relations of 'proper' plasma waves, for which k is real.

The solutions for which ω is also real will occur at points S such that:

$$\left.\begin{array}{c} f_0'(\omega/k)=0 \\ \mathrm{P.}\displaystyle\int_{-\infty}^{\infty}\frac{f_0'(u)}{u-\omega/k}\,du \geqslant 0 \end{array}\right\} \qquad \ldots\ldots(18)$$

A suitable point S, found for any distribution, corresponds to $\omega/k\to\infty$, where according to (17)

$$\omega=\omega_0, \qquad k=0. \qquad \ldots\ldots(19)$$

The points S are at the intersection with the real axis of the dispersion curves defined by (16) in the complex plane ω/k, along which k is real (for more details see Berz 1955, p. 30).

The characteristics of the solutions depend on the function f_0. It can be shown (Berz 1955, pp. 39, 82) that for any given real value of k, (i) amplified or undamped waves, either do not arise or are finite in number, (ii) damped waves always exist, in finite or in infinite number.

2.3. *Case of Isotropic Distributions of the Maxwellian Type*

Of particular importance are the isotropic distributions $f_0(u)$ such that

$$f_0(u)=f_0(-u) \qquad \ldots\ldots(20\,a)$$

$$f_0'(u)<0 \ \text{ for } \ 0<u<\infty \qquad \ldots\ldots(20\,b)$$

$$f_0'(0)=f_0'(\infty)=0. \qquad \ldots\ldots(20\,c)$$

For that class of functions all the plasma waves are damped, except the wave given by (19)(see Appendix III).

This is true in particular for the Maxwellian distribution,† in which case, moreover, for any given k there is an infinity of damped solutions ω (Berz 1955, p. 82). For such a distribution equations (16) take the form

$$\left(\frac{\omega}{\omega_0}\right)^2 = -2z^2 - 2iz^3\sqrt{\pi}(1+\operatorname{erf} iz) \quad \ldots\ldots(21)$$

where $z = \omega/k(2\overline{u^2})^{\frac{1}{2}}$.

An approximate solution of (21) for the wave of smallest damping can be obtained by means of a first order expansion with respect to $\mathscr{I}\omega$ (Berz 1955, p. 52). Figure 1 gives the results.

For wave velocities $V_w > 5(\overline{u^2})^{1/2}$, not represented on the figure, the damping is very small, due to the small number of electrons travelling at these velocities. Equation (17) is satisfied with a good approximation, and whereas the frequency ω varies only very slowly, $\omega_0 < \omega < 1{\cdot}05\omega_0$, the wavelength λ covers a very wide range $30\lambda_D < \lambda < \infty$. For gas discharges the Debye length λ_D is of the order of 1 mm or below

$$\lambda_D = \left(\frac{kT}{4\pi ne^2}\right)^{1/2} \simeq \left(48\frac{T}{n}\right)^{1/2} \text{ cm} \quad (T = \text{electron temperature in °K}).$$

Therefore although these waves have a dispersion equation, the result found by Langmuir when neglecting the velocity distribution, i.e. $\omega = \omega_0$, λ arbitrary, gives a fair approximation.

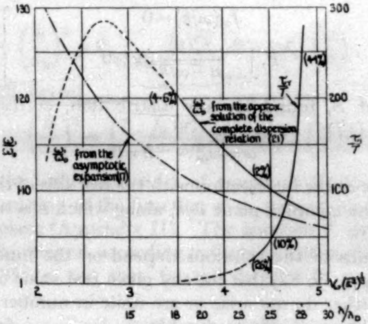


Figure 1. Plasma oscillation of smallest damping excited by ordinary initial disturbances, when the equilibrium velocity distribution is Maxwellian.

The figures in brackets give the maximum relative error introduced by the approximate method of solution of the dispersion equation (21). These errors become important in the region where the curves are dotted. V_w = wave velocity, $(\overline{u^2})^{1/2}$ = r.m.s. electron velocity in the direction of the wave propagation; ω_1/ω_0 = (frequency of oscillation)/(plasma frequency); ω_1 is the real part of ω; T = period of oscillation, $T = 2\pi/\omega_1$; τ_r = ' relaxation time ', i.e. the interval of time after which the square of the field amplitude is decreased by the factor e.

---

† It is assumed in this section that the distribution extends to $u = \infty$. This however does not introduce any substantial difference as will be seen from the remarks in § 3.4.

As shown in figure 1, the damping increases when ω is increased. It becomes important for $\omega > 1{\cdot}15\omega_0$, $V_w < 4(\overline{u^2})^{1/2}$. At the same time the dispersion relation between the real parts of ω and k deviates markedly from (17).

2.4. *Conclusions*

It must be noted that although the results of this section have been derived for given initial conditions at $t=0$, they are also likely to apply in substance when initial conditions are given at $x=0$.

This means that there are 'proper' plasma waves, whose dispersion characteristics are independent of the condition of excitation.

For Maxwellian distributions these waves are damped, except for the solution (19). The real parts of ω and k are asymptotically related by (17) for large values of ω/k. The damping becomes important in the immediate neighbourhood of the characteristic frequency ω_0.

§ 3. SOME REMARKS ABOUT THE DISPERSION EQUATION OF VLASSOV, AND BOHM AND GROSS

3.1. *The Equation and its Limitations*

The dispersion equation considered in the important papers of Vlassov (1945), and Bohm and Gross (1949 a, b) is of the form

$$\left(\frac{k}{\omega_c}\right)^2 = \int_{-\infty}^{\infty} \frac{f_0'(u)}{u-\omega/k}\,du. \qquad \ldots\ldots(22)$$

This equation is derived by means of a 'substitution analysis', i.e. by assuming that $\tilde{f}$ as well as $\tilde{E}$ varies in space and time as $e^{i(kx-\omega t)}$. This amounts to taking $C(u)=0$ in equation (8).

The lack of generality of this procedure pointed out by Landau (1946) and Twiss (1952), can be understood more fully in view of the preceding sections. As seen in §§ 1 and 2, equation (22) has general validity for amplified waves only. Besides this limitation, the methods used for the solution of (22) are also open to criticism as seen in what follows.

3.2. *Interpretation of the Equation for Real Values of ω and k*

The interpretation of (22) for ω/k real is somewhat delicate. Although the integral in (22) diverges, when substituting real values of ω/k, for $f_0'(\omega/k) \neq 0$,† it tends towards finite limits, when the real axis of ω/k is approached through paths in the complex ω/k plane (Appendix II). To assume that these limits give the actual value of k^2/ω_0^2 for real values of ω/k means that there is no discontinuity in the wave propagation when ω/k is real, and when it has a small imaginary component. We can thus write for (22):

For $\mathscr{I}\,\omega/k \neq 0$

$$\left(\frac{k}{\omega_0}\right)^2 = \int_{-\infty}^{\infty} \frac{f_0'(u)}{u-\omega/k}\,du. \qquad \ldots\ldots(23\,a)$$

---

† The divergence of the integral is due to the electrons which travel near or at the wave velocity ω/k, and the influence of which is greatly over-stressed by the linear treatment, leading to (22). In actual fact these electrons have but a finite effect.

For $\mathscr{I}\omega/k=0$

$$\left.\begin{aligned}\left(\frac{k}{\omega_0}\right)^2 &= \lim_{|\epsilon|\to 0}\int_{-\infty}^{\infty}\frac{f_0'(u)}{u-\omega/k-i|\epsilon|}\,du\\ &= \mathrm{P.}\int_{-\infty}^{\infty}\frac{f_0'(u)}{u-\omega/k}\,du+i\pi f_0'\left(\frac{\omega}{k}\right)\end{aligned}\right\}\quad\ldots\ldots(23\,b)$$

or

$$\left.\begin{aligned}\left(\frac{k}{\omega_0}\right)^2 &= \lim_{|\epsilon|\to 0}\int_{-\infty}^{\infty}\frac{f_0'(u)}{u-\omega/k+i|\epsilon|}\,du\\ &= \mathrm{P.}\int_{-\infty}^{\infty}\frac{f_0'(u)}{u-\omega/k}\,du-i\pi f_0'\left(\frac{\omega}{k}\right)\end{aligned}\right\}\quad\ldots\ldots(23\,c)$$

The values of the limits given by (23 *b*, *c*) are justified in Appendix II, and are valid for infinite and finite distributions ($f_0(u)\equiv 0$ for $|u|>u_{\mathrm{M}}$).

As seen by comparison with (16 *b*), equation (23 *b*) corresponds to a 'proper' plasma wave, whereas (23 *c*) is obtained only with special conditions of excitation.

3.3. *Discussion of the Solution*

Consider now the solution of equation (23) for the case of isotropic distributions, $f_0(u)=f_0(-u)$. In the papers quoted in § 3.1 use is made of (17). Neglecting the second term on the right-hand side of (23 *b*, *c*), the following dispersion equation is obtained

$$\left(\frac{\omega}{\omega_0}\right)^2=1+3\left(\frac{k}{\omega}\right)^2\overline{u^2}.\qquad\ldots\ldots(24)$$

It would follow from this equation that for $\omega\geqslant\omega_0$, real values ω/k correspond to real values of k and ω. But these results are not correct, due to the errors introduced by the approximations leading from equations (23) to (24). The rigorous discussion of the existence of solutions must be made along the line used for (16 *a*, *b*) in § 2.2.

It can be seen that the solutions of (23) with k and ω real occur only at points defined by (18). For continuous and finite distributions, $f_0(u)\equiv 0$ for $|u|>u_{\mathrm{M}}$, this includes solutions with $\omega/k>u_{\mathrm{M}}$. If moreover $(\overline{u^2})^{1/2}/u_{\mathrm{M}}\ll 1$ such solutions satisfy (24) and therefore $\omega\simeq\omega_0$, which is very similar to the result of Langmuir, with the added restriction $\omega/k>u_{\mathrm{M}}$.

For the important class of finite isotropic distributions such that $f_0'(u)<0$ for $0<u<u_{\mathrm{M}}$, $f_0'(u)\equiv 0$ for $u\geqslant u_{\mathrm{M}}$, no other solutions of (23) can occur with either ω or k real, except for $\omega<\omega_0$ and k imaginary† (see Appendix III).

3.4. *Conclusions*

A comparison with the results of §§ 1 and 2 shows that the dispersion equation (22) lacks generality, and that its solutions are often restricted to values of k such that $\omega/k>u_{\mathrm{M}}$. This occurs because of the rather narrow conditions put *a priori* on $\bar{f}$. When no such conditions are imposed, as in the initial boundary problem of § 2, solutions arise for all values of k. It is interesting to note that for infinite Maxwellian distributions the solutions of § 2 for $V_{\mathrm{w}}>5(\overline{u^2})^{1/2}$ are very similar to those found in the present section when $f_0(u)\equiv 0$ for $|u|\geqslant 5(\overline{u^2})^{1/2}$.

† In connection with the finite distributions I would like to acknowledge a private communication from P. C. Clemmow and A. J. Willson, where a natural cut-off to electron velocities is introduced by a relativity treatment. *Note added in proof.* Their work is to appear in *Proc. Roy. Soc.* A, 1956, **237**.

§ 4. General Practical Conclusions

It is seen that the plasma waves can be divided into two groups.

In the first group are the damped waves which depend on the conditions of excitation.

In the second are 'proper' plasma waves, which are determined by the stationary conditions of the plasma, and which can be either amplified, damped, or progressive (with no amplification or damping), according to the type of the velocity distribution function f_0. Progressive plasma waves can only be found for values of ω/k such that $f_0'(\omega/k)=0$.

For Maxwellian distributions the 'proper' plasma waves are progressive ($\omega=\omega_0$) or damped. As shown in figure 1 the damping per period of oscillation is small for $1\leqslant\omega/\omega_0<1\cdot05$, $30\lambda_D<\lambda<\infty$ and in this range the real parts of ω and k are asymptotically related by equation (17). For $\omega/\omega_0>1\cdot15$ the damping per period is very strong.

Acknowledgments

The investigation described in this paper was carried out in the Electrical Engineering Department of the Imperial College. The author wishes to thank Dr. D. Gabor for his most helpful interest and encouragement, and to express her great indebtedness to Mullard Research Laboratories for financial support.

APPENDIX I

The Use of Laplace Transforms for given Initial Conditions

The electric field and velocity distributions can be put in the forms

$$\tilde{E}=\mathscr{E}(t)\,e^{ikx}$$

$$\tilde{f}=F(u,t)\,e^{ikx},$$

with $F(u,0)=g(u)$. Let $\mathscr{E}_p$ and $f_p=F_p(u)\,e^{ikx}$ be the corresponding Laplace transforms with respect to time.

Equation (3) gives

$$F_p=\frac{1}{p+iku}\left[g(u)+\frac{e}{m}\mathscr{E}_p\frac{df_0}{du}\right]$$

whence it follows, using (4), that

$$\mathscr{E}_p=i\frac{4\pi ne}{k}\frac{G(p)}{1-I(p)} \qquad \ldots\ldots(25)$$

where for $\mathscr{R}p>0$

$$G(p)=\int_{-\infty}^{\infty}\frac{g(u)}{p+iku}\,du \qquad \ldots\ldots(26)$$

$$I(p)=i\frac{\omega_0^2}{k}\int_{-\infty}^{\infty}\frac{f_0'(u)}{p+iku}\,du. \qquad \ldots\ldots(27)$$

For $\mathscr{R}p\leqslant0$, $G(p)$ and $I(p)$ are the analytic continuations of the functions defined by the integrals (26), (27) for $\mathscr{R}p>0$. These are derived according to the principles developed in Appendix II.

The electric field $\mathscr{E}(t)$ is obtained from (25) by means of the inversion integral. This can be evaluated using the residues which arise at the poles of $G(p)$ and at the roots of the equation $1-I(p)=0$.

Hence the results of § 2.

APPENDIX II

The Integral $\int_{-\infty}^{\infty} \frac{f_0'(u)}{u-\omega/k}\,du$.

It is first assumed that $f_0'(u)$ is analytic with

$$\int_{-\infty}^{\infty} |f_0'(u)|\,du$$

finite. Landau (1946) has shown that under these conditions the functions of ω/k defined by the sums on the right-hand sides of (16 *b*) and (16 *c*) give the analytic continuation of the function given by the integral on the right-hand side of (16 *a*).† In view of its importance this result is derived here, by a slightly different method.

Let C be a contour in the lower half-plane u such that $f_0'(u)$ has no singularities either on C or between C and the real axis (see figure 2 (*a*) where $u_1, u_2, \ldots$ are the singularities of $f_0'(u)$). The function $F_+(\omega/k)$ is defined, for points ω/k above C, by the integral

$$F_+\left(\frac{\omega}{k}\right) = \int_C \frac{f_0'(u)}{u-\omega/k}\,du. \qquad \ldots\ldots(28)$$

According to a well-known theorem (Copson 1935, p. 107) $F_+(\omega/k)$ is analytic and without singularities above C.

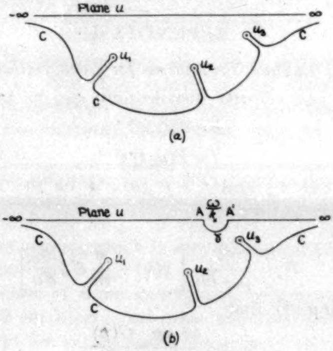


Figure 2

Applying the theorem of residues to the contour formed by C and the real axis (figure 2 (*a*)) equation (28) can be put in the following equivalent form:

For $\mathscr{I}\omega/k > 0$

$$F_+\left(\frac{\omega}{k}\right) = \int_{-\infty}^{\infty} \frac{f_0'(u)}{u-\omega/k}\,du$$

and for ω/k between C and the real axis

$$F_+\left(\frac{\omega}{k}\right) = \int_{-\infty}^{\infty} \frac{f_0'(u)}{u-\omega/k}\,du + 2\pi i f_0'\left(\frac{\omega}{k}\right).$$

---

† This result can be also derived from Plemelj formulae (see Muskhelishvili 1953), with less restrictions on $f_0'(u)$, which must satisfy certain continuity conditions, but is not necessarily analytic.

For $\mathscr{I}\omega/k=0$ Cauchy's theorem is applied along the contour $\mathrm{C}\infty\mathrm{A}'\gamma\mathrm{A}-\infty$ (see figure 2 (*b*)), giving

$$F_+\left(\frac{\omega}{k}\right)=\int_{-\infty}^{\infty}\frac{f_0'(u)}{u-\omega/k}\,du+\pi i f_0'\left(\frac{\omega}{k}\right).$$

Since the contour C is arbitrary, this proves the result stated above concerning equations (16). In these formulae the integrals along the real axis define functions of ω/k without any singularities (Copson 1935), and the only singularities of $F_+(\omega/k)$ are those of f_0' in the lower half-plane.

It could be shown in a similar way that equation (23 *b*) gives the analytic continuation of (23 *a*). This also holds in the case of finite distributions, such that $f_0'(u)$ is analytic for $|u|<u_{\mathrm{M}}$, and is equal to zero for $|u|>u_{\mathrm{M}}$. (In this case $f_0'(\omega/k)=0$ in (23 *b*) for $\omega/k>u_{\mathrm{M}}$.) The derivation is the same as above with the contour C of figure 2 terminating at the points $u=\pm u_{\mathrm{M}}$.

Similarly the last line on the right-hand side of (23 *c*) gives the analytic continuation on the real axis of the function $F_-(\omega/k)$, defined for $\mathscr{I}\omega/k<0$ by

$$F_-\left(\frac{\omega}{k}\right)=\int_{-\infty}^{\infty}\frac{f_0'(u)}{u-\omega/k}\,du.$$

APPENDIX III

On the Absence of Undamped Solutions for Certain Types of Isotropic Distributions

First consider solutions of (16 *a*, *b*) with k real for distributions of the type defined by equation (20). Let us put $\omega/k=X+iY$. The only possible solution along $Y=0$, i.e. for both ω and k real, must satisfy (18), and is therefore given by equation (19).

Outside the real axis (16 *a*) takes the form

$$\begin{aligned}\left(\frac{k}{\omega_0}\right)^2&=\int_{-\infty}^{\infty}\frac{f_0'(u)}{u-X-iY}\,du\\&=\int_{-\infty}^{\infty}\frac{(u-X)f_0'(u)}{(u-X)^2+Y^2}\,du+iY\int_0^{\infty}\left[\frac{1}{(u-X)^2+Y^2}-\frac{1}{(u+X)^2+Y^2}\right]f_0'(u)\,du.\end{aligned}\qquad\ldots\ldots(29)$$

For $X\neq0$, $Y\neq0$, the second integral is always negative, and k cannot be real.

For $X=0$ (29) can be written in the form

$$\left(\frac{\omega}{\omega_0}\right)^2=-2Y^2\int_0^{\infty}\frac{uf_0'(u)}{u^2+Y^2}\,du.\qquad\ldots\ldots(30)$$

Thus along $X=0$, ω is real and k is imaginary.

Hence for k real no undamped waves can arise, besides the solution (19).

Multiplying both sides of (29) by $(\omega/k)^2$ it could similarly be shown that besides the solution given by (19) ω is real only along the axis $X=0$, where k is imaginary and $\omega<\omega_0$, as follows from the inequality

$$-2Y^2\int_0^{\infty}\frac{uf_0'(u)}{u^2+Y^2}\,du<-2\int_0^{\infty}uf_0'(u)\,du$$

and from

$$-2\int_0^{\infty}uf_0'(u)\,du=2\int_0^{\infty}f_0(u)\,du=1.$$

The results can be easily extended to the solutions of (23) for isotropic finite and continuous distributions $f_0(u)$ such that $f_0'(u)<0$ for $0<|u|<u_M$, and $f_0'(u)\equiv 0$ for $|u|\geqslant u_M$. It can be seen that in that case the only solutions with k and ω real are those given in § 3.3.

REFERENCES

BERZ, F., 1955, *Thesis*, University of London.
BOHM, D., and GROSS, E. P., 1949 a, *Phys. Rev.*, **75,** 1851; 1949 b, *Ibid.*, **75,** 1864.
COPSON, E. T., 1935, *The Theory of Functions of a Complex Variable* (Oxford : Clarendon Press), p. 107.
ECKER, J., 1955 a, *Z. Phys.*, **140,** 274; 1955 b, *Ibid.*, **140,** 293; 1955 c, *Ibid.*, **141,** 294.
LANDAU, L., 1946, *J. Phys. U.S.S.R.*, **10,** 25.
LANGMUIR, I., and TONKS, L., 1929, *Phys. Rev.*, **33,** 195.
MUSKHELISHVILI, N. I., 1953, *Singular Integral Equations* (Groningen : Noordhoff).
VLASSOV, A., 1945, *J. Phys. U.S.S.R.*, **9,** 25.
TWISS, R., 1952, *Phys. Rev.*, **88,** 1392.
VLASSOV, A., 1945, *J. Phys. U.S.S.R.*, **9,** 25.
WHITTAKER, E. T., and WATSON, G. N., 1927, *A Course of Modern Analysis* (Cambridge : University Press), p. 115.

RESEARCH NOTES

The Measurement of the Electron Content of the Ionosphere by the Lunar Radio Echo Method

By J. V. EVANS

Jodrell Bank Experimental Station, University of Manchester

Communicated by A. C. B. Lovell ; MS. received 11th June 1956

The investigation of the long period fading of radio echoes reflected from the moon (Murray and Hargreaves 1954, Browne, Evans, Hargreaves and Murray 1956) indicated a possible new technique for the measurement of the total ionospheric electron content. This note describes the preliminary results of such measurements. The method consists of measuring the amount of rotation of the plane of polarization of the radio waves during their double passage between the earth and the moon. It can be shown that for the most likely distribution of electrons between the earth and moon, about 98% of the rotation will occur in the region 100 to 500 km. Thus equation (2*a*) of Browne *et al.* (1956) can be written as

$$\omega = 3{\cdot}35 \times 10^{-12}\overline{B\cos\theta\sec\chi}\int_{100}^{500} N\,dh \text{ radians} \qquad \ldots\ldots(1)$$

where the numerical constant includes $\boldsymbol{e}$, $\boldsymbol{m}$, $\boldsymbol{c}$ and the radio frequency f, and ω = rotation of the plane of polarization (radians), B = terrestrial magnetic induction (gauss), θ = angle between the line of sight and the magnetic lines of force, χ = angle between the line of sight and the zenith at height h.

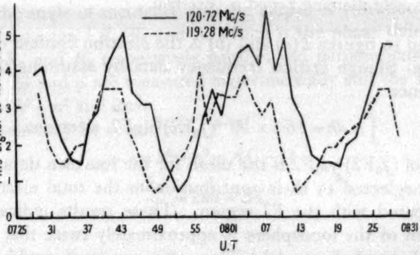


Figure 1. The r.m.s. signal-to-noise ratio on two frequencies on 9th November 1955.

The rotation ω has been measured by observing the change in amplitude of radio echoes from the moon received on a horizontally polarized aerial array. The frequency of the equipment is switched every 25 pulses (three quarters of a minute) between 119·28 Mc/s and 120·72 Mc/s. The signal-to-noise ratio on the two frequencies shows a similar variation, but there is a time displacement between the curves. An example of this is shown in figure 1 and the derivation of ω from such curves has been described previously (Browne *et al.* 1956).

The aerial system restricts the use of the present equipment to a period of ± 30 minutes about lunar transit on 14 days per lunation and this Note refers to the lunations of October and November 1955. Each day's run commenced approximately where the previous one finished, thus the period midnight to 10.0 a.m. was covered. Because of these limitations a single day's run cannot provide an unambiguous result. For example, the minima in figure 1 were found to be $3\frac{1}{4}$, $3\frac{3}{4}$ and $4\frac{1}{4}$ rotations respectively, with an uncertainty such that they could be all $+\frac{1}{2}$ rotation or $-\frac{1}{2}$ rotation. The ambiguity can be eliminated by joining the curves obtained on successive days. If the ionosphere behaves in the same manner from day to day, then continuity would be expected and this has been found to be the case with the exception of some early morning runs in November. Figures 2(*a*) and (*b*) show the variation of the electron content $\int N\,dh$ from $h = 100$ to 500 km, deduced from the curves for these periods.

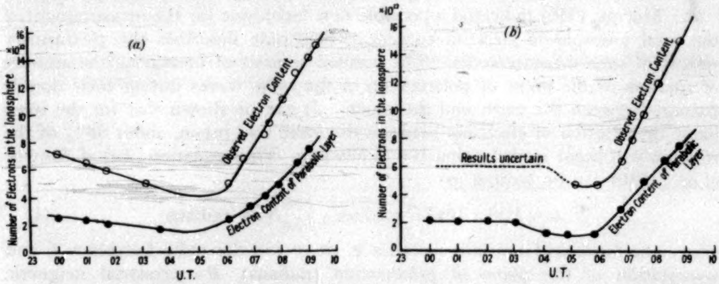


Figure 2. The variation of the measured electron content, together with that estimated from Slough critical frequency data. The ordinates give the number of electrons in the ionosphere per cm² along the line of sight. (*a*) October 1955; (*b*) November 1955; the early morning period is uncertain.

Also plotted in figures 2(*a*) and (*b*) is the electron content of the ionosphere calculated from Slough critical frequency data by assuming that the region is parabolic, whence

$$\int N\,dh = 16{\cdot}5 \times 10^{-9} (f_0\mathrm{F}2)^2 y_m\mathrm{F}2 \text{ electrons.} \qquad \ldots\ldots(2)$$

The value of $(f_0\mathrm{F}2)^2 y_m\mathrm{F}2$ is the mean for the fourteen days. The E and F1 regions were neglected as their contribution to the total electron content will be small compared with the F2 region. These results indicate that the total electron content of the ionosphere is approximately twice that expected on the basis of a simple parabolic model. Thus, after sunrise it would appear that there are three times as many electrons above the level of the maximum as below, whilst before sunrise this ratio is somewhat higher. A similar result for the distribution of electrons in the F region was obtained by Smith (1952) from his studies of the refraction of radio waves from radio stars on 81·5 Mc/s. From the work of Chapman (1931) a ratio of about 2:1 would be expected. Gledhill and Szendrei (1950), who have extended Chapman's work to the case of a non-isothermal atmosphere, show that this ratio depends critically upon the temperature gradient and is always greater than that predicted by Chapman when this is positive.

REFERENCES

BROWNE, I. C., EVANS, J. V., HARGREAVES, J. K., and MURRAY, W. A. S., 1956, *Proc. Phys. Soc.* B, **69**, 901.

CHAPMAN, S., 1931, *Proc. Phys. Soc.* B, **43**, 26, 483.

GLEDHILL, J. A., and SZENDREI, M. E., 1950, *Proc. Phys. Soc.* B, **63**, 427.

MURRAY, W. A. S., and HARGREAVES, J. K., 1954, *Nature, Lond.*, **173**, 944.

SMITH, F. G., 1952, *J. Atmos. Terr. Phys.*, **2**, 350.

Tables for quick Calculation of Optical Constants and Variation of the Ellipse of Vibration with Angle of Incidence

BY P. CROSBY†

West Ham College of Technology, London E.15

MS. received 17*th August* 1955, *and in revised form* 7*th March* 1956

§ 1. CALCULATION OF OPTICAL CONSTANTS FROM ELLIPSE OF REFLECTION

IN 1946 Price derived exact equations for the calculation of the optical constants of a specimen from the ellipse of vibration obtained when plane-polarized light of 45° azimuth is reflected obliquely by a plane surface of the specimen. His equations were:

$$-\epsilon = k^2 - n^2 = \frac{2\sin^2\phi\tan^2\phi\,(\cos\Delta+\sin 2\psi)}{\sin 2\psi\,(\cos\Delta+\mathrm{cosec}\,2\psi)^2} - \tan^2\phi \qquad \ldots\ldots(1)$$

$$\frac{2\sigma}{\nu} = 2nk = \frac{2\sin^2\phi\tan^2\phi\sin\Delta}{\tan 2\psi\,(\cos\Delta+\mathrm{cosec}\,2\psi)^2} \qquad \ldots\ldots(2)$$

where n is the refractive index, k the extinction coefficient, ϵ the dielectric constant and σ the electrical conductivity of the specimen at the frequency ν of the reflected radiation. ϕ is the angle of incidence and the ellipse of vibration is specified by the azimuth angle $\psi = \tan^{-1} R_{\mathrm{p}}/R_{\mathrm{s}}$ and Δ, which is the phase difference between the components R_{p} and R_{s} respectively parallel and perpendicular to the plane of incidence. Δ, 2ψ and ϕ are measured experimentally and the Price equations are solved to find k^2-n^2 and $2nk$.

These equations may be rewritten:

$$-\epsilon = k^2 - n^2 = AC - D \qquad \ldots\ldots(3)$$

$$\frac{2\sigma}{\nu} = 2nk = BC \qquad \ldots\ldots(4)$$

where

$$A = \frac{(\cos\Delta+\sin 2\psi)}{\sin 2\psi\,(\cos\Delta+\mathrm{cosec}\,2\psi)^2}, \qquad B = \frac{\sin\Delta}{\tan 2\psi\,(\cos\Delta+\mathrm{cosec}\,2\psi)^2},$$

$$C = 2\sin^2\phi\tan^2\phi, \qquad D = \tan^2\phi.$$

The values of C and D are easily determined. A and B are complicated and have been tabulated to four significant figures for all combinations of Δ and 2ψ at degree intervals in the range $0 < \Delta < 180°$ and $45° < 2\psi < 90°$.

---

† Now at Weapons Research Establishment, Salisbury, South Australia.

Hence the greater part of the calculation has been carried out and the solution of the Price equations has been reduced to the simple arithmetic of calculating $AC-D$ and BC in order to determine k^2-n^2 and $2nk$ respectively. Calculations that would have taken several days to work through now only require an hour or so. The tables are kept in the Physical Society archives and at the Science Library.

§ 2. CALCULATION OF ELLIPSE OF VIBRATION FROM OPTICAL CONSTANTS

Other uses of the tables have been described elsewhere (Crosby 1955). The most interesting is the reverse calculation in which the ellipse of vibration at any angle of incidence may be deduced from a knowledge of the optical constants. So far it has not been possible to produce a pair of equations for Δ and 2ψ in terms of k^2-n^2, $2nk$ and ϕ, but the tables enable the calculations to be carried out, thus:

Rewriting equations (3) and (4), one obtains

$$A=\frac{k^2-n^2+D}{C}=\frac{k^2-n^2+\tan^2\phi}{2\sin^2\phi\tan^2\phi} \qquad \ldots\ldots(5)$$

$$B=\frac{2nk}{C}=\frac{2nk}{2\sin^2\phi\tan^2\phi}. \qquad \ldots\ldots(6)$$

The values of k^2-n^2 and $2nk$ being known, A and B are calculated for a given angle of incidence ϕ. The new tables are then scanned to find the combination of Δ and 2ψ which give these values of A and B simultaneously. Experience has shown that the combination of Δ and 2ψ may be found to within $\frac{1}{4}$ degree of arc in each angle.

By calculating the ellipse of vibration for various angles of incidence one may determine the family of ellipses for all angles of incidence.

REFERENCES

CROSBY, P., 1955, *Ph.D. Thesis*, University of London.
PRICE, D. J., 1946, *Proc. Phys. Soc.*, **58**, 704.

Spectroscopic Evidence for Presence of Atoms in Active Nitrogen

BY D. T. STEWART
Physics Department, Queen's University of Belfast

Communicated by K. G. Emeléus; MS. received 3rd April 1956

ALTHOUGH most recent theories of the Lewis–Rayleigh afterglow of nitrogen agree in associating it with volume recombination of nitrogen atoms, there is relatively little direct spectroscopic evidence for their presence (cf. Herzberg 1953, Herzfeld and Broida 1956). The afterglow would not be expected to emit visible or infra-red dipole lines of the NI spectrum, as these require at least 11·7 ev for excitation from the ^{4}S ground state, and indirect spectroscopic evidence indicates that the maximum energy available for excitation is only 9·8 ev (Mitra 1945). Nitrogen possesses, however, low metastable ^{2}D and ^{2}P states, with excitation potentials of 2·4 v and 3·6 v respectively, and it seemed possible that the 'forbidden' radiative transitions between these states, or between one or both and the ground state might be detectable.

Actually conditions for observation of the lines are less favourable than for observation of the corresponding auroral and transauroral lines of OI (Sayers and Emeléus 1952). The ^2D–^2P multiplet of NI has wavelengths *c.* 10400 Å, and is therefore liable to be obscured by first positive bands of N_2. The ^4S–^2D transition at 5200 Å has a total Einstein *A* coefficient of only about $3{\cdot}5 \times 10^{-5}$ sec^{-1} (Pasternack 1940), which makes it doubtful if it can be observed at all in spontaneous emission from laboratory sources. That leaves the ^4S–^2P doublet at 3466·4 Å and 3466·7 Å, with *A* coefficients of $6{\cdot}7 \times 10^{-3}$ and $2{\cdot}7 \times 10^{-3}$ sec^{-1} respectively. This has been produced from other laboratory sources, for example, by Janin (1946) and R. and L. Herman (1946), and attention was therefore concentrated on searching for it in the light from the afterglow.

The afterglow was generated in streaming gas in a Pyrex tube 5 cm in diameter and 100 cm long, with a quartz window at one end. Commerical nitrogen from a cylinder containing less than 1% of oxygen was passed through pyrogallol, and dried by tubes containing phosphorus pentoxide and traps cooled by liquid oxygen. After activation in a heavy Tesla spark discharge, it entered the observation tube through a side arm near the quartz window. The speed of the gas in the observation tube was of the order of 50 cm sec^{-1}. At pressures in the range of 1 to 50 mm Hg, the observation tube was filled with the yellow afterglow, and provided a source of considerable intensity.

The spectrum was photographed through the quartz end, using an f/5 quartz prism spectrograph with a dispersion of 85 Å mm^{-1} at 3670 Å. An isolated line with wavelength $3466{\cdot}5 \pm 0{\cdot}3$ Å was recorded at pressures of from 2 to 10 mm Hg, with exposure times of from 2 to 4 hours. Examination of the infra-red, visible and ultra-violet regions revealed in addition only the 2537 Å line of HgI, and some bands of N_2 and NO. The intensity distribution in the first positive bands was that typical of active nitrogen (Reinecke 1953), and there was no correlation between the intensity of the line at 3466 Å and the intensities of the NO bands. It was concluded that the line was the unresolved doublet of NI, and that N atoms in the ^2P state are present in the afterglow. Presumably N atoms in the ^2D state and the ^4S ground state are present in greater numbers. From observations on the production of infra-red dipole lines of NI by electron collisions with N_2 molecules (Stewart 1955), it is thought likely that the N atoms in the three low states (^2P, ^2D and ^4S) are largely produced in the parent discharge by dissociative ionization.

ACKNOWLEDGMENT

It is a pleasure to acknowledge the interest taken by Professor K. G. Emeléus in this work.

REFERENCES

HERMAN, R., and HERMAN, L., 1946, *J. Phys. Radium*, **7**, 203.
HERZBERG, G., 1953, *Disc. Faraday Soc.*, **14**, 127.
HERZFELD, C. M., and BROIDA, H. P., 1956, *Phys. Rev.*, **101**, 606.
JANIN, J., 1946, *Thesis*, University of Paris.
MITRA, S. K., 1945, *Active Nitrogen* (Calcutta : Indian Press).
PASTERNACK, S., 1940, *Astrophys. J.*, **92**, 129.
REINECKE, L. H., 1953, *Z. Phys.*, **135**, 361.
SAYERS, N. D., and EMELÉUS, K. G., 1952, *Proc. Phys. Soc.* A, **65**, 219.
STEWART, D. T., 1955, *Proc. Phys. Soc.* A, **68**, 404.

On the Anomalous Rectification of Cuprous Sulphide Detectors

By M. ANASTASSIADES and D. ILIAS
University of Athens

MS. received 10*th May* 1955, *and in revised form* 20*th February* 1956

The direction of rectification of a metal semiconductor detector is, as is well known, given by the semiconductor type (n or p). There are some rare cases, where the rectification for increasing a.c. potentials changes from p to n or conversely. Typical examples of this effect are given by metal contacts on crystal covellite, on crystal chalcosine, or on a disc of pure CuS powder (Anastassiades 1933).

In this latter case the direction of rectification for low potentials is metal to semiconductor and is reversed for potentials higher than 0·3 volt r.m.s. Thus the detector under discussion rectifies as p or n in accordance with the applied potential (figure 1).

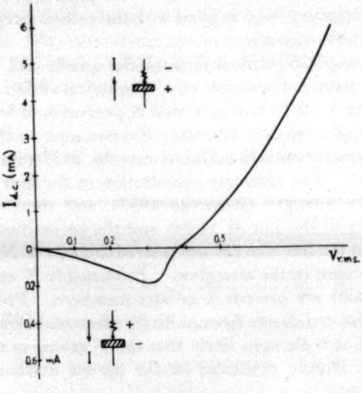


Figure 1.

To explain the above phenomena the variation of the response of the CuS detector with frequency for potentials higher and lower than the reversing potential was studied by an oscillographic method. A typical oscillogram of a nearly static characteristic of the system taken for a frequency of a few cycles and a potential greater than 0·3 volt r.m.s. is shown in figure 2. The branch OBC corresponds to a negative point with a positive semiconductor. The shape of the nearly static characteristic, which does not show any hysteresis loop, changes substantially when a potential of the order of 0·3 volt is exceeded. Hysteresis loops and parts of negative resistance appear connecting distinctive branches of the system OA to OD and OB to OC.

It is found that the slope of the branch OC depends on the frequency. Hysteresis loops appear on both branches for very low values of frequency (period 10–30 sec). For increasing values of frequency the area of the loop changes asymmetrically. It may here be observed that of the two loops that of the branch ODA having the polarity point semiconductor diminishes more rapidly in area; for frequencies of 250 c/s and above it is practically non-existent. Simultaneously and for the reversed polarity the loop area diminishes while the breakdown potential of B is as much delayed as the frequency increases. This loop also disappears for frequencies of the order of 20 000 c/s (figure 3).

Figure 2.

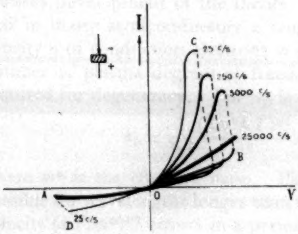


Figure 3.

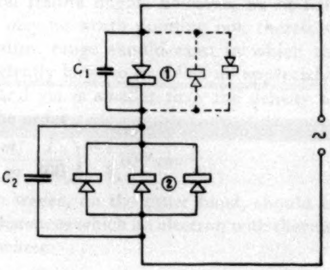


Figure 4.

To explain the mechanism of inversion of detection in CuS it may be considered that there are two contacts, the point and the holder, in series and in opposition. This leads to an equivalent circuit consisting of a rectifier in series and opposition with a group of several other rectifiers of the same type in parallel bypassed by a large condenser. This group represents the holder action (figure 4). It can be assumed that the holder consists of several points in parallel and is consequently of low total resistance. It can also be assumed, because of the large area presented, that capacity effects appear. An analogous circuit was used by Nikreschevich and Fisher (1949) to explain point–PbS detector action.

Thus the oscillogram of figure 2 may be easily explained as follows: for a potential up to 0·3 volt the system is conductive in the direction point to crystal (branches OA and OB). Above 0·3 volt the point contact breaks down losing its rectification abilities and becoming a low resistance. The system now rectifies by means of the crystal holder contact (branches OC and OD). A constant differential action exists between these contacts, and their oscillogram branches are combined by regions of negative resistance. The order of time duration of this negative resistance which was determined to be 10^{-5} second coincides with the disappearance of the hysteresis loop at frequencies of the order of 20 000 c/s.

REFERENCES

ANASTASSIADES, M., 1933, *C.R. Acad. Sci., Paris*, **197**, 677, 1397.
NIKRESCHEVICH, and FISHER, 1949, *J. Tech. Phys. U.R.S.S.*, **19**, 1312.

LETTERS TO THE EDITOR

Plasma Interaction and Conduction in Semiconductors

The scattering of electrons by lattice vibrations, which is responsible for the temperature dependent part of the electric resistivity, is, in metals, restricted to the relatively few electrons in a range of the order kT near the top of the Fermi surface. The existence of plasma oscillations has only little direct influence on this scattering, though indirectly of course it influences the parameters determining the scattering probability. In non-degenerate semiconductors, on the other hand, the usual treatment assumes that all electrons are subject to scattering by the lattice vibrations (apart from other types of scattering). The influence of plasma interactions should be noticeable here much more directly than in metals, because the total number of degrees of freedom involved in plasma interactions will here, in general, be temperature dependent. Hence it should be expected that the establishment of plasma vibrations should alter the $T^{-3/2}$ law for the mobility due to lattice scattering; the greater the number of plasma degrees of freedom, the larger the mobility should be. For, if the plasma would completely describe the electronic motion, then no ordinary scattering at all would take place.

A satisfactory theoretical treatment of this effect requires a much clearer understanding of the dynamics of the electrons than that provided by current plasma theory. Analysis of experimental results might, however, be of help towards development of the theory. It may be worth pointing out, therefore, that in many semiconductors a temperature range should exist in which the density n of conduction electrons is sufficiently large to establish an appreciable number of plasma degrees of freedom and yet is smaller than the density n_D required for degeneracy. For n_D is of the order

$$n_D \simeq \frac{2(2\pi m^* kT)^{3/2}}{h^3} \sim \left(\frac{m^*}{m}\frac{T}{300}\right)^{3/2}\frac{1}{4}10^{20}\,\mathrm{cm}^{-3},$$

where m^* is the effective mass. Plasma waves, on the other hand, should be possible for wavelengths longer than the distances which an electron with thermal velocity $(kT/m^*)^{1/2}$ covers in a period ω where

$$\omega^2 = \frac{4\pi}{\epsilon}\frac{ne^2}{m^*}$$

and ϵ = dielectric constant.

The number of plasma modes is large when this length becomes of the order $n^{-1/3}$. This takes place when n is of the order, or larger than, n_p, where

$$= \left(\frac{\epsilon}{4\pi}\frac{kT}{e^2}\right)^3 \simeq \left(\frac{\epsilon}{4\pi}\frac{T}{300}\right)^3 \frac{1}{2}10^{16}\,\mathrm{cm}^{-3}.$$

Clearly $n_D \gg n_P$ is often feasible.

Department of Theoretical Physics,
University of Liverpool.
9th July 1956.

H. FRÖHLICH.
S. DONIACH.

REVIEWS OF BOOKS

Static and Dynamic Electron Optics, by P. A. STURROCK. Pp. x + 240. Cambridge Monographs on Mechanics and Applied Mathematics (Cambridge: University Press, 1955.) 30*s*.

Among the many treatises on electron optics this is a book apart. It bears on every page the imprint of the author's hand, who is a manipulative mathematician of rare fluency, with a strong *penchant* for elegance, order and conciseness. Without conciseness it would have been impossible to collect such a vast amount of material between the covers of a small-octavo book of only 240 pages, but many a reader might wish that another 100 pages or so could have been added for explaining the results in terms more familiar to the average physicist.

Within the given limitations of space, the author has done extremely well in covering not only ordinary ' static ' electron optics, in a wider frame than usual, and from a higher point of view, but also ' dynamic electron optics ', that is to say the theory of particle accelerators. This last third of the book is probably its most important part, and fills a real gap in textbook literature.

The salient feature of this work is the systematic use of variational principles, with Lagrangians (which the author regrettably calls ' variational functions '), expanded in series of perturbation characteristics of progressive order. Almost every result is obtained by this powerful method, often by admirably short cuts, and many of these results are the author's own. The deductive process naturally starts from the top of the pyramid, and descends towards the basis. It brings with it, unavoidably, a didactical difficulty. A. and L. Föppl, who deduced the whole of elasticity theory in a similar way from Castiglione's Principle illustrated it by saying that one starts with a million-pound note, which has to be converted gradually into smaller change. The average reader will not understand much of it until the author has arrived at the small change to which he is used. Indeed, only few physicists, with a very mathematical turn of mind, can be expected to learn any chapter of physics by a descent from the top of a pyramid. For the others it will be necessary to read Sturrock's chapters first forwards, then backwards, and finally forwards again. In the end most of them will learn that the repeated journey was not only necessary, but the best way to prepare them for original work.

Other features which make the reading rather difficult are the somewhat unorthodox notations, such as e.g. the ' ray equation ' (2.2.10),

$$d\mathbf{p}/ds = \partial n/\partial \mathbf{x}.$$

This is admirably concise, but the reader might be excused if he does not know what is meant by division by a vector. The unorthodox definitions are scattered over the pages and physical concepts are too often referred to by their symbol instead of by name. Clarendon does not always mean a vector in the book, **p** stands for the vector momentum as well as for the scalar value of the momentum on the axis. But in most cases the reader will have to admit that Dr. Sturrock's notations, as well as his units, are more logical and practical than those in common usage.

This book will be indispensable to all those who want to do original work in electron optics or on particle accelerators, and it must be seriously recommended to all mathematically minded physicists who have acquired an elementary knowledge of electron optics, and want an introduction to its higher flights.

D. GABOR.

Questions actuelles en luminescence cristalline, par Maurice Curie et Daniel Curie. Pp. 86. (Paris: Editions de la Revue d'Optique Théoretique et Instrumentale, 1956.) 600 fr.

This brief survey might be described as the thinking aloud of two well-known workers in this field about the facts which seem to them now specially significant and about some of the interpretations.

The three principal parts—*Structure of a luminescent sulphide*, *Displacement of electrons and holes in the crystal lattice* and *Transitions without emission of light*, deal chiefly with the luminescence and photoconductivity of zinc and cadmium sulphides with some side glances at other phosphors. There is a short final section on *Cathodoluminescence*. The authors are concerned with the contribution that the study of luminescence and associated electrical effects can make to the knowledge of the solid state. In a field in which experimental results provide abundant material for many theories yet defy integration in one, there is bound to be, in a short survey, exclusions which some will find surprising. For example, in the most valuable section dealing with the basic problem of efficiency it is a pity that some room could not have been found for a discussion of Dexter and Schulman's ideas and for the implications of the conception of a large centre. However the treatment is generally remarkably unbiased and clear; generous and constructive in its criticisms of current theories. It is no introductory treatise for beginners hot for certainties, for the answers are seen to be dusty ones, but it is welcome as mildly provocative review of some current facts and fancies.

J. EWLES.

Analysis of Deformation. Vol. II, *Experiment and Applied Theory*, by K. Swainger. Pp. xxxvi + 365. (London: Chapman and Hall, 1954.) 70*s*.

This is the second volume of a three volume work, two of which have so far appeared. It is a well-produced book, albeit an expensive one. Since most of the material of the second volume depends entirely upon the validity of the *Mathematical Theory* of Vol. I, which has been found very much wanting, it is difficult to place much reliance on the interpretation of such experimental work as is considered.

As with Vol. I, one can hardly say it is easy to read, even for the professional mathematician, much less for those to whom it is addressed. The dust jacket expresses the hope that the book will supply " a need for the research engineer, experimentalists and physicists without unduly emphasising the mathematics from which the subject cannot be divorced ". The subject cannot indeed be divorced from sound mathematical foundations and it is precisely for this reason that the reviewer is unable to recommend this second volume any more than the first. The author speaks of a Vol. III in preparation which considers elasto-viscous fluids, renders boundary layer methods unnecessary and reveals deficiencies in the classical theory of hydrodynamics. These claims are in line with those of Vol. I, where however the alleged deficiencies in the classical theory of elasticity arise from elementary errors in the mathematical development of the foundations of the subject by the author, commented upon at length by the present reviewer in his review of Vol. I in this journal (*Proceedings of the Physical Society* B, 1954, **67**, 596).

A. C. STEVENSON.

CONTENTS OF SECTION A

PAGE

Dr. B. L. Moiseiwitsch. Interaction Energy and Charge Exchange between Helium Atoms and Ions 653

Prof. H. S. W. Massey and Mr. R. O. Ridley. Application of Variational Methods to the Theory of the Scattering of Slow Electrons by Hydrogen Molecules . 659

Dr. S. Yoshida. The Inelastic Scattering of Nucleons by the Surface Interaction 668

Dr. G. W. Greenlees and Dr. A. E. Souch. The Scattering of 9·6 mev Protons by Mg 686

Dr. M. H. Alston, Dr. A. V. Crewe, Dr. W. H. Evans and Dr. G. von Gierke. A Study of Positive Pion Production in p–p Collisions at 383 mev 691

Dr. K. W. Allen, Dr. B. Collinge, Dr. B. Hird, Mr. B. C. Maglić and Dr. P. R. Orman. The Proton Gamma Angular Correlations in the ^{28}Si(d, pγ)^{29}Si Stripping Reaction 705

Research Notes :

Dr. R. W. Nicholls, Mr. D. Robinson, Mr. W. Parkinson and Mr. W. R. Jarmain. r-Centroids for the O_2^+ Second Negative, N_2 Lyman–Birge–Hopfield, and N_2 Vegard–Kaplan Band Systems 713

Reviews of Books 716

Contents of Section B 724

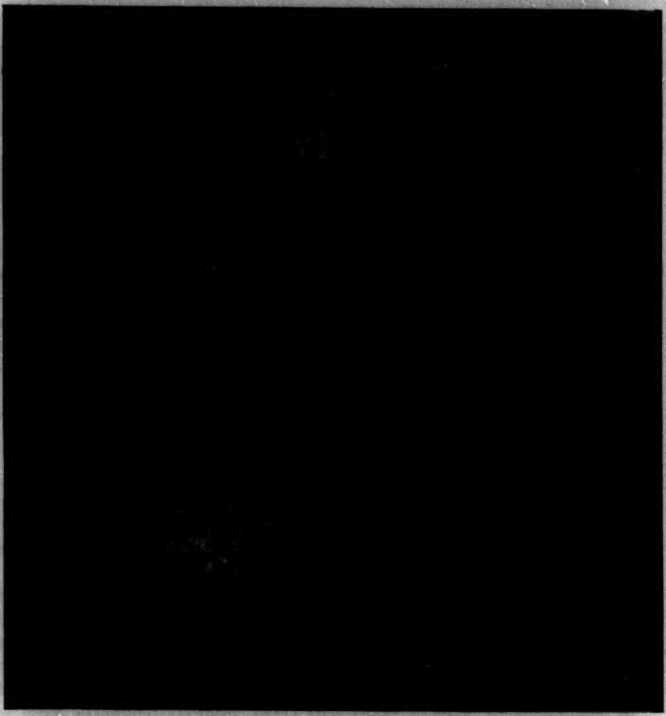


Figure 1. Kyropoulos boule.

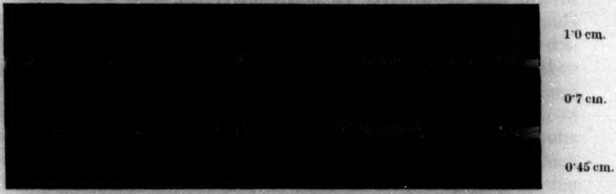


Figure 5. Photograph showing variation of [100] in (001) for lengths of irradiated strip 1·0, 0·7 and 0·45 cm (in [010] direction).

Measurements of the Velocity and Absorption of High-Frequency Ultrasonic Waves in Supercooled Liquids

BY A. N. HUNTER
Physics Department, University College of Leicester

MS. received 10th February 1956, and in revised form 28th March 1956

Abstract. Measurements of ultrasonic velocity and absorption have been made in five polar liquids, including the supercooled region. Discontinuities occur in the gradient of the velocity–temperature relation similar to those observed by others in the case of viscosity and dielectric constant. The absorption rises steeply below the melting point and minima occur above the melting point. Simple theories based on a suspension of 'crystallites' are shown to be inadequate. The fall in the ratio of observed to shear viscosity absorption in the supercooled region is suggestive of association.

§ 1. INTRODUCTION

THE purpose of this paper is to present experimental results for the absorption and velocity of ultrasonic waves in five supercooled liquids at frequencies in the neighbourhood of 70 Mc/s and to discuss some implications of the effects observed. The five liquids are diphenyl ether, salol, menthol, azoxybenzene and benzophenone. All are polar and fairly readily supercooled and the first four were selected because the accurate viscosity measurements necessary to the calculations were available from the work of Dodd and Hu Pak Mi (1949).

It has been explained (Hunter 1951) that these acoustic measurements in supercooled liquids were suggested by the work of Dodd and Hu Pak Mi (1949) on the variation of viscosity with temperature in diphenyl ether: since their paper, a number of measurements have been made of various physical properties of supercooled liquids and these are summarized by Greenwood and Martin (1952). In brief it is found that certain physical properties of polar liquids (possibly those which are sensitive to molecular weight) exhibit a marked discontinuity in gradient at the melting point when the liquid sample is taken through the nominal melting point without freezing. The effect is small and can only readily be demonstrated when there is a known linear dependence of the property under investigation with temperature. If the experimental results are used to plot such a relation they are found to fit two straight lines which intersect close to the melting point. Dodd and Hu Pak Mi (1949) consider that this intersection denotes the melting point, although Greenwood and Martin (1952) find that the transition region is not sharp. It has not been firmly established that the effect is peculiar to polar liquids and no such phenomenon has been detected experimentally in the case of density and surface tension. The present work adds ultrasonic velocity to the list of properties which show the discontinuity and also indicates that further work on the variation of ultrasonic absorption with frequency in supercooled liquids may throw light on the mechanisms involved.

§ 2. REQUIREMENTS OF SUPERCOOLED LIQUIDS

The following points are of importance to the measurements in supercooled liquids:

1. Since stirring of the liquid makes supercooling impossible the apparatus must be able to maintain a uniform temperature throughout a static liquid sample for long periods.
2. A small bulk of liquid increases the chances of supercooling (de Coppet 1907) and also assists the attainment of a uniform temperature and of chemical cleanliness.
3. Metal parts should be avoided where possible since some limit seriously the amount of supercooling. (Copper was found to be bad in this respect although gold, aluminium and stainless steel seemed less detrimental.)
4. Because of the smallness of the effect sought, the apparatus should be designed with a view to selection of temperatures at random in order to avoid systematic errors.
5. Supercooling is more certain if the sample is taken to a high temperature after re-melting (the highest temperature used was 110°c). Too rapid a cooling appears to limit the amount of supercooling possible.
6. A high rate of working is desirable so that changes in apparatus parameters during a 'run' are small and so that a wide range of liquids and frequencies may be explored in a reasonable time.

The apparatus used was particularly troublesome with respect to the last point. Although a measuring run could be obtained in about 30 minutes a period of several hours was normally allowed for thermal equilibrium to be established and this limited the number of points obtained per day to one or two.

§ 3. EXPERIMENTAL METHOD

3.1. *General*

The measurements described have all been made by the now familiar pulse technique (Pellam and Galt 1946, Pinkerton 1949 b) and only details of particular interest are given here. Most of the measurements were at a nominal frequency of 70 Mc/s (representing the seventh harmonic of a 10 Mc/s crystal) and a double crystal method was used in order to simplify the electrical design and reduce overall loss. Absorption measurement was facilitated by pulsing the transmitter and signal generator on alternate cycles of the time base. In this way two pulses were superposed on the screen of the cathode-ray tube, so simplifying adjustment to equality: in other respects the method of making absorption measurements was that of Pinkerton (1949 b). In the case of wavelength (and hence velocity) determination the two sources were simultaneously pulsed, the signal generator pulse being made much wider than the acoustic pulse, and the two signals being made coherent by injection of a small locking signal from the transmitter into the signal generator. The frequencies were adjusted to equality by tuning for zero beat within the overlapping region of the two pulses, the output from the amplifier being rectified before display, and the changes of phase of the smaller acoustic pulse as the acoustic path was varied appeared as an alternate swinging above and below the flat top of the wider superposed signal generator pulse. In general at least 150 'bobs' of the pulse could be counted, corresponding to a movement of 150 wavelengths, and at the higher temperatures the number was limited by the extent of the liquid rather than by absorption.

A bar of fused quartz with optically plane ends was used as an 'acoustic delay line' (Rapuano 1947) and also as a support for the experimental vessel. One feature of the method was the use of the unavoidable connecting cables as resonant elements in the matching network thus preventing their appearance as shunt capacitances with consequent loss of bandwidth.

3.2. *The Interferometer Cell*

As seen in figure 1 the fused quartz bar A was mounted vertically through the base of a bath containing thermostatically controlled oil, the joint being sealed by a neoprene rubber ring, and the bar was further supported by a clamp attached to the same base but not shown in the figure. The experimental cell B was of glass and attached to A with a ground joint, the transmitting crystal Q_2 being attached to the exposed lower end face of A with a film of oil. Much difficulty was experienced at first in making the joint between A and B proof against diffusion of oil from the bath into the sample liquid. (No 'grease' which was not soluble in the hot organic liquids used could be found for the joint, and the various gaskets tried instead of the ground joint all gave the same trouble.) The problem was solved by lubricating the joint with the experimental liquid itself and providing the skirt C: this skirt trapped a layer of air beneath it so as to give a liquid–air seal.

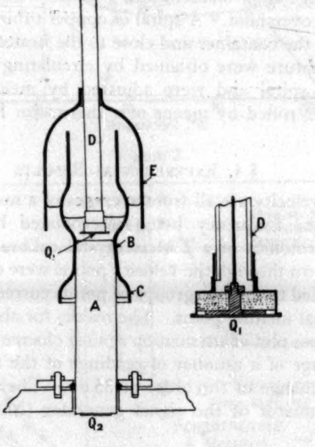


Figure 1.

The receiving crystal Q_1 was soldered to the assembly at the end of the glass supporting tube D. This supporting tube was silvered internally so as to form part of the coaxial line system while still preserving a low thermal conductivity.

It was also found necessary to enclose the sample completely by providing the 'diving bell' E, the assembly being submerged beneath the oil. Until this was done the sample liquid was found to differ in temperature from that of the bath by amounts up to ½°c or so. The stirrer in the oil bath was of paddle type and supported independently so as to avoid communicating vibration to the acoustic parts. The whole lower part of the apparatus rested on a ball and cup bearing, the centre of the spherical surface coinciding approximately with the centre of the upper end face of A. By means of this device the lower crystal could be adjusted for parallelism with the upper crystal without change in mean distance, the adjustment being made acoustically: details of the design of the bearing will be published elsewhere.

The temperature of the liquid was determined by using a thermocouple as a differential thermometer to indicate any difference in temperature from that of the surrounding bath, the temperature of the bath itself being measured with a certified mercury-in-glass thermometer. The bath was heated by a spiral filament wound round the inner circumference of the container in order to provide even heating: this filament was of bare wire to obtain good thermal contact and reduce overshoot. A spiral of copper tubing was also wound round the inner surface of the container and close to the heater winding: temperatures below room temperature were obtained by circulating refrigerated methylated spirit through this spiral and were adjusted by means of the heater. The temperature was controlled by means of a thermistor bridge.

§ 4. EXPERIMENTAL RESULTS

The results for velocity are all from averages of a number of counts of about 200 wavelengths, the frequency being interpolated by the signal generator (checked against harmonics of a 2 Mc/s crystal calibrator). The equations to the straight lines drawn through the velocity points were calculated by the method of least squares applied to the two groups of points corresponding to values above and below the normal melting point. The results for absorption are from a least squares slope of a linear plot of attenuation against change in liquid path or, in some cases, from an average of a number of readings of the movement necessary for a fixed attenuation change of the order of 35 dB. The standard of attenuation was the piston attenuator of the signal generator (Marconi Ekco TF517/F1) accurate to about 2%.

Figures 2 and 3 show the variation of velocity and absorption with temperature for two of the five liquids investigated: for convenience the absorption has been expressed in dB mm^{-1}. These two liquids have been selected as being representative of the results obtained and all the liquids show a nearly linear variation of velocity with temperature in both the supercooled and normal states. According to Rao (1940) the velocity of sound raised to the power 10/9 should be proportional to the temperature and the velocity points have also been tested against this empirical law. The inset diagrams show the deviations of all experimental points from the least squares line for Rao's law applied to the points above the melting point and thus accentuate the discontinuity in slope. The velocity results for all five liquids are expressed in the following equations. The equations are those to the two straight lines corresponding to values above and below the melting point respectively, for each liquid.

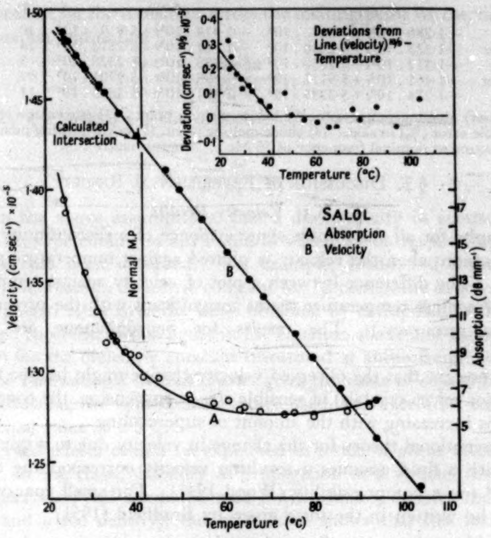


Figure 2

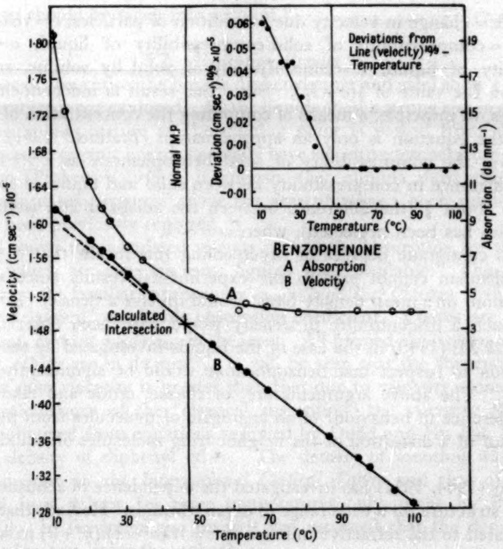


Figure 3

	(1)	(2)	(3)	(4)	(5)
Salol	$-1\cdot280\,.\,10^3t+5\cdot864\,.\,10^5$	$-1\cdot414\,.\,10^3t+5\cdot919\,.\,10^5$	10	0·46	1·1
Benzophenone	$-1\cdot323\,.\,10^3t+6\cdot2038\,.\,10^5$	$-1\cdot502\,.\,10^3t+6\cdot2872\,.\,10^5$	18	0·062	0·3
Azoxybenzene	$-1\cdot312\,.\,10^3t+6\cdot2050\,.\,10^5$	$-1\cdot385\,.\,10^3t+6\cdot2328\,.\,10^5$	5	0·15	1·26
Diphenyl ether	$-1\cdot461\,.\,10^3t+5\cdot9117\,.\,10^5$	$-1\cdot598\,.\,10^3t+5\cdot9701\,.\,10^5$	8·6	0·42	0·7
Menthol	$-1\cdot274\,.\,10^3t+5\cdot3235\,.\,10^5$	$-1\cdot421\,.\,10^3t+5\cdot3865\,.\,10^5$	11	0·07	2·1

(1) $c^{10/9}$ above melting point; (2) $c^{10/9}$ below melting point; (3) percentage change in slope; (4), (5) probable error (%) in slope: (4) above melting point, (5) below melting point. c(cm sec^{-1}) = velocity of sound at nominal frequency of 70 Mc/s; t = temperature (°c).

§ 5. DISCUSSION OF EXPERIMENTAL RESULTS

5.1. *Velocity Results*

The graphs for all the liquids show evidence of a discontinuity of the type under discussion when the velocity is plotted against temperature according to Rao's law. (The difference between a plot of velocity against temperature and (velocity)$^{10/9}$ against temperature seems insignificant with the present degree of experimental accuracy.) The results for benzophenone are particularly convincing.

It was thought that the observed velocity change might be due to molecular aggregates (or micro-crystals) of sensible size in suspension, the concentration of these groups increasing with the amount of supercooling.

The conventional theory for the change in velocity due to a mixture of solid particles with a fluid assumes a resulting velocity corresponding to the mean density and mean compressibility (Wood 1941). For small concentrations the result may be written in the form given by Bradfield (1951)

$$\frac{\delta c}{c} = -(\tfrac{1}{2}\sigma + \tfrac{1}{2}K - 1)C$$

in which δc = change in velocity due to addition of particles, c = velocity in pure liquid, K = compressibility of solid/compressibility of liquid, σ = density of solid/density of liquid, C = concentration of solid by volume, and the sign depends on the value of $\tfrac{1}{2}(\sigma + K)$. Since this result is independent of particle size it gives, in principle, a means of computing the concentration of solid phase, although the equation is only an approximation (Bradfield 1951). Bridgman (1915) gives the compressibility of solid benzophenone as $2\cdot9 \times 10^{-5}\,\mathrm{cm}^2\,\mathrm{kg}^{-1}$ and the difference in compressibility between solid and liquid as about 1×10^{-5} thus K is about $\tfrac{1}{3}$ (the difference between the adiabatic and isothermal compressibilities has been neglected), whereas $\sigma \sim 1$.

For 20 centigrade degrees of supercooling this means that $C \sim 3\cdot6 \times 10^{-2}$. This mechanism cannot explain the experimental results since such a high concentration, on a mean density basis, would involve a density change of about 0·1%. Such a discontinuity in density would have been observed by Dodd and Hu Pak Mi (1949) in the case of the liquids investigated by them and there is no reason to suspect that benzophenone would be significantly different in behaviour. The above arguments are, of course, crude and take no account of any difference in behaviour of an aggregate of molecules from that of a solid particle, nor of a distortion of the neighbouring molecules of liquid by such an aggregate.

Schaafs (1944, 1951) has investigated the dependence of acoustic velocity on molecular structure for a wide range of organic liquids. He finds that the velocity can be related to the refractive index n by $c = W(n^2 - 1)/(n^2 + 2)$ in which W can be treated as constant for a great variety of liquids and the value of n should strictly be that corresponding to the acoustic frequency. If we now *assume* that W is a constant for the transition across the melting point we can differentiate the above equation and obtain

$$\frac{\delta c}{c}=\frac{6n\,\delta n}{(n^2+2)(n^2-1)}.$$

Replacing n by $\sqrt{\epsilon}$, where ϵ is the dielectric constant at the same frequency we obtain

$$\frac{\delta c}{c}=\frac{3\,\delta\epsilon}{(\epsilon+2)(\epsilon-1)}.$$

It follows, on the above assumptions, that a discontinuity in acoustic velocity should occur on supercooling in any liquid in which the corresponding dielectric constant effect has been observed. For example, Dodd and Roberts (1950) find about 3% increase in dielectric constant for 12 centigrade degrees of supercooling in diphenyl ether over the value obtained by extrapolation from higher temperatures. Insertion of this figure in the equation above and using the single value of 3·(9) for the dielectric constant (measured at audio frequencies) which is given in the *International Critical Tables* gives the value of $\delta c/c$ to be expected as 2%. Considering the approximate values used, this agrees remarkably with the experimental value of 2·5%.

The observed effects cannot be explained in terms of shear viscosity alone. For instance, in the classical theory of propagation in a viscous medium, it is usual to neglect $4\omega^2\eta^2/9\rho^2$ compared with c^4 (in which η denotes the coefficient of viscosity and ρ the density), but calculation shows that this introduces an error of about 1 part in 10^5 in the present measurements and would cause a fall in velocity at the lower temperatures instead of a rise. (At still higher frequencies this correction should be taken into account.)

5.2. *Absorption Results*

In the case of absorption the liquids diphenyl ether and salol show a minimum in the temperature range explored. All the liquids show a strong rise in absorption with decreasing temperature as the melting point is passed from above and, since unassociated organic liquids normally have a positive temperature coefficient of absorption (Pinkerton 1949 a), it appears that minima must exist in other cases although at temperatures above the present range: such minima appear not to have been previously reported.

Figure 4 shows the calculated viscous part of the absorption for those three liquids for which accurate viscosity figures were available. This has been computed from the familiar expression $\alpha'_{\text{class}}=2\eta\omega^3/3\rho c^3$ (Stokes 1845) in which α'_{class} is the 'classical' amplitude absorption coefficient. Curves are also given for $\alpha_{\text{obs}}-\alpha_{\text{class}}$, and for $\alpha_{\text{obs}}/\alpha_{\text{class}}$ in the inset. All primed absorption coefficients are in cm^{-1}, others are in dB mm^{-1}. It will be seen that the absorption due to causes other than viscosity is greater than that due to viscosity alone.

In the calculations of the 'classical absorption' the viscosity of all three liquids was obtained from equations supplied by Dodd and Hu Pak Mi (1949), as was the density of diphenyl ether. The density of menthol was obtained from the equation in the *International Critical Tables* and that of salol was extrapolated from values, for three temperatures only, given by Burriel and Marti (1930). In the latter two cases it was assumed that the density versus temperature law was linear throughout.

An attempt was made to explain the experimental results for absorption by means of the 'particle suspension' theory already discussed in connection with the velocity change. The increased absorption due to the addition to a fluid of small rigid incompressible spheres has been theoretically investigated by Lamb (1945, p. 361) and Urick (1948). The calculation results in a scattering term proportional to the fourth power of the frequency and a term proportional to the frequency, due to shear viscosity. Both terms are functions of particle size and concentration. The maximum contribution of both terms has been calculated by taking as the maximum radius of an aggregate the thermodynamical critical radius (of the order 10^{-5} to 10^{-6}) above which crystallization is certain. The result gives a maximum of 0·5 dB mm^{-1} for the scattering and 0·05 dB mm^{-1} for the viscosity term: both at 70 Mc/s and the high volume concentration of 0·1%. Although insignificant at this frequency the scattering could, on this basis, be appreciable at rather higher frequencies, since a fourth power law is involved. However the absence of optical scattering in supercooled liquids would seem to rule out the possibility of acoustic scattering at frequencies below 1000 Mc/s.

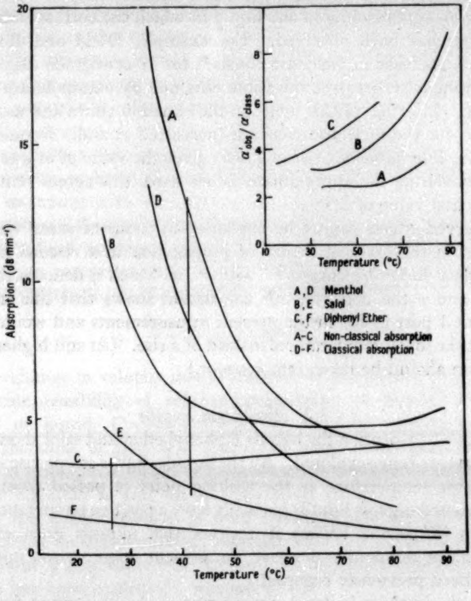


Figure 4

In order to check that the contribution due to scattering was negligible an experiment was carried out with menthol at 68·0 Mc/s and at 87·5 Mc/s. These two frequencies are rather close together but a wider range could not be obtained without extensive alterations to the apparatus; the results are given in the following table.

Ratio of Absorptions at 87·5 Mc/s and 68·0 Mc/s

Temperature (°c)	34	36	38	40	42	45	50	55	60	65
α_{obs}	1·60	1·66	1·63	1·64	1·64	1·65	1·64	1·62	1·55	1·55
$\alpha_{\mathrm{obs}}-\alpha_{\mathrm{class}}$	1·58	1·66	1·63	1·63	1·63	1·62	1·65	1·60	1·49	1·51

These figures were obtained from smooth curves drawn through the experimental points corresponding to each frequency. The square of the ratio of the two frequencies is 1·654 and the fourth power is 2·73. It will be seen that both the total absorption and the 'non-classical' part alone give good agreement with the first value and show that any scattering is negligible.

Returning to figure 4, it will be noted that the curves for the 'non-classical' part of the absorption have the same general shape as those for the total observed absorption. This is to be expected since the classical part decreases smoothly with increase in temperature.

It has been suggested (E. R. Andrew, private communication) that the experimental results of Dodd and Hu Pak Mi (1949) and of Greenwood and Martin (1952) may be explained in terms of a small progressive association of molecules, in which dimers would probably predominate, as the supercooling proceeds. The plots given in the inset to figure 4 show that the ratio of the observed to classical absorption falls with decreasing temperature until in the neighbourhood of the melting point it is about 2·5. Pinkerton (1949a) has noted that associated polyatomic liquids have values of this ratio all lying between 1 and 3 whereas non-associated polyatomic liquids give values between 3 and 1500. Thus the experimental curves give evidence for a change from an unassociated type of liquid to an associated type in the supercooled state: that is, a progressive association with lowering of temperature.

§ 6. CONCLUSION

The experiments described in this paper show that the number of liquid properties which exhibit the so-called discontinuity at the melting point may be extended to include the acoustic velocity. There is a complementary effect on the absorption, which however cannot be displayed in the same way as the velocity effect as there is no linear function to plot.

In the case of a relaxation process the maximum dispersion $\delta c_{\max}$ is related to the maximum absorption per wavelength $\mu_{\max}$ by $\mu_{\max}/\pi=\delta c_{\max}/c$ and Bradfield (1951) has shown that the same relation is true for the case of solid particles suspended in a fluid. In the case of benzophenone at 20° of supercooling $\mu c/\delta c \sim 3$ in which δc simply refers to the increase of velocity over that extrapolated from higher temperatures and μ to the non-classical part of the absorption. Although structural relaxation may be a factor, only experiments over a wide range of frequency will show if this is significant.

The objection that the method of displaying the velocity results will tend to exaggerate any change in slope at the melting point applies also to the work of others on different physical properties, although both Dodd and Hu Pak Mi (1949) and Greenwood and Martin (1952) have used statistical methods to justify this procedure. The results obtained in the present investigation give some evidence for a progressive association as the temperature falls through the melting point and also for the belief that any liquid which shows the effect with dielectric constant should also show it in the case of acoustic velocity. Further work over a wide frequency range should yield useful information about the mechanisms involved, but there would seem to be no good reason for duplicating the temperature variation measurements. The only other published measurements of acoustic effects in supercooled liquids appear to be those of Litovitz (1951) in glycerine and some recent measurements of acoustic velocity in palmitic acid by Gabrielli and Verdini (1955): the latter found no discontinuity in velocity in this liquid.

ACKNOWLEDGMENTS

The work described in this paper has been the subject of a Ph.D. Thesis for the University of London, and earlier experiments were commenced with the aid of the Dixon Fund. The author wishes to thank Professor E. A. Stewardson for the facilities to continue the work, Professor H. T. Flint for encouragement and advice and Professor E. R. Andrew for many valuable suggestions.

REFERENCES

BRADFIELD, G., 1951, *Overdruck uit het Colloquium over Ultrasonore Trillingen.* (Brussels: Koninklijke Vlaamse Academie voor Wetenschappen).

BRIDGMAN, P., 1915, *Phys. Rev.*, **6**, 1, 94.

BURRIEL-MARTI, F., 1930, *Bull. Soc. Chim. Belg.*, **39**, 590.

DE COPPET, L., 1907, *Ann. Chim.* (*Phys.*) **8**, ser. 10, 457.

DODD, C., and HU PAK MI, 1949, *Proc. Phys. Soc.* B, **62**, 454.

DODD, C., and ROBERTS, G., 1950, *Proc. Phys. Soc.* B, **63**, 814.

FRENKEL, J., 1946, *Kinetic Theory of Liquids* (Oxford : Clarendon Press).

GABRIELLI, I., and VERDINI, L., 1955, *Ric. Scientifica*, **25**, 1125.

GREENWOOD, N., and MARTIN, R., 1952, *Proc. Roy. Soc.* A, **215**, 46.

HUNTER, A. N., 1951, *Proc. Phys. Soc.* B, **64**, 1086.

LAMB, H., 1945, *Hydrodynamics*, 6th Edn. (Cambridge : University Press).

LITOVITZ, T., 1951, *J. Acoust. Soc. Amer.*, **23**, 75.

PELLAM, J., and GALT, J., 1946, *J. Chem. Phys.*, **14**, 608.

PINKERTON, J., 1949 a, *Proc. Phys. Soc.* B, **62**, 129; 1949 b, *Ibid.*, **62**, 286.

RAO, R., 1940, *Ind. J. Phys.*, **14**, 109.

RAPUANO, R., 1947, *Phys. Rev.*, **72**, 78.

SCHAAFS, W., 1944, *Z. Phys. Chem.*, **194**, 28; 1951, *Nuovo Cim.*, **7**, Supplement 2, ser. **9**, 286.

STOKES, G., 1845, *Trans. Camb. Phil. Soc.*, **8**, 287.

URICK, R., 1948, *J. Acoust. Soc. Amer.*, **20**, 283.

WOOD, A., 1941, *Textbook of Sound* (London : Bell), p. 361.

Changes in Conductivity resulting from Breakdown in Cadmium Sulphide Single Crystals

By J. WOODS

Research Laboratories of The General Electric Company Limited, Wembley, Middx.

MS. received 29*th February* 1956, *and in revised form* 4*th May* 1956

Abstract. For small applied fields the resistivity of a pure cadmium sulphide crystal is very high. For fields greater than a critical value, however, dielectric breakdown occurs and large currents flow. If the crystal is protected from destruction by a limiting resistor it is found on re-measurement at small fields, after the passage of a high current, that the conductivity and photosensitivity have been greatly increased. The increase in conductivity is greater, the larger the current passed through the crystal after the initial breakdown. This activating effect has been examined in detail, and it is shown that the changes which occur as a result of the passage of high currents take place largely in the bulk of the crystal.

§ 1. Introduction

Dielectric breakdown effects in cadmium sulphide single crystals have been described by Boer and Kümmel (1952), and Boer, Kümmel and Rompe (1952). From their experiments they concluded that as a steadily increasing direct electric field is applied to the crystal, breakdown occurs either by a field or thermal ionization process. Field breakdown usually occurs first for insulating crystals in the dark or under weak illumination, whereas for crystals under stronger illumination thermal breakdown takes place first.

Diemer (1954) has also investigated the breakdown of CdS crystals. He demonstrated that certain features of the breakdown are analogous to processes occurring in gas discharges. Thus the initial breakdown process of an insulating crystal in the dark is analogous to the Townsend process. Once large currents have begun to flow, however, the temperature of the crystal rises and the breakdown becomes largely thermal in character.

The work to be described here is substantially in agreement with that of Diemer, although the range of currents has been extended and some different features observed. In particular more attention has been given to the irreversible changes in electrical properties of the crystal, which occur at various stages of the breakdown.

§ 2. Experimental

The single crystals used were grown by the Frerichs (1947) technique and were of the usual thin plate-like variety. No activation treatment was given. These crystals were only moderately photoconducting with resistivities of 10^{10}–10^{12} ohm cm in the dark which fell to 10^{8}–10^{10} ohm cm under illumination from a 60 watt tungsten filament lamp, one foot away.

Indium contacts were used in most experiments, although silver and graphite electrodes were also used. The contacts were applied in such a way that the current flow was parallel to the largest dimension of the crystal.

§ 3. The First Breakdown Process

With the crystal in the dark, an increasing direct voltage was applied to a series combination of specimen and load resistor. At small voltages the current was small and increased linearly with voltage. However, at a critical field strength of about 5×10^4 volts cm^{-1} the current increased rapidly. A graph of current flow plotted against voltage across the crystal is shown in figure 1, curve A. This reveals that after breakdown had occurred a negative resistance characteristic was described.

At the onset of breakdown, flashes of red luminescence were observed which were soon extinguished. After breakdown, however, most crystals exhibited a green luminescence which was localized into a few fine streaks parallel to the striations which are a feature of the surface topography of the crystals. Provided currents in excess of 0·5 mA were not passed, the green luminescence could be maintained indefinitely. With currents exceeding 0·5 mA the green luminescence disappeared and was replaced by a red emission, which in its turn was soon extinguished. If the current was subsequently reduced, no luminescence could be detected. These results contrast sharply with those of Boer and Kümmel and of Diemer who observed a red luminescence localized at the anode.

Since the optical absorption edge of CdS shifts to longer wavelengths with increasing temperature, the body colour of a crystal can be used to indicate its temperature. It was found that the temperature rose to a maximum and then fell again as the negative resistance part of the characteristic was described. The negative slope is due, in part, to the increased electrical conductivity which results from the increase in temperature.

If after curve A, figure 1, had been obtained the voltage was reduced to zero and then increased again, curve B was obtained. Further repetition of this sequence led to curve C. A curve such as B was not displaced towards C if the current was kept below 100μA.

The extent of the displacement of the current–voltage characteristics to lower voltages was dependent on the length of time which elapsed before the new characteristic was determined. Thus if the new characteristic, B say, was measured immediately after curve A had been obtained, the maximum excursion to lower voltages was obtained. However, if several hours were allowed to elapse before the new characteristic was measured, a curve intermediate between A and B was obtained. The intermediate characteristic did not revert completely to the position of the original curve, even after several days. This shows that the total change which occurs in a crystal as a result of passing a large current is comprised of a permanent and a non-permanent part.

By controlling the current flow, crystals were obtained with dark resistivities permanently reduced to values within the range 10^{10} to 1 ohm cm. Temporary changes in resistivity of the type discussed in the previous paragraph were only observed with crystals whose post-breakdown dark resistivities were in the range 10^{10}–10^6 ohm cm. The photoconductive properties of a crystal were also changed by the breakdown and maximum photosensitivity was obtained with crystals with a dark resistivity of 10^8 ohm cm. This resistivity could be reduced to 10^3–10^4 ohm cm on illumination with a 60 watt tungsten lamp.

The temporary change in conductivity is presumably due to a redistribution of electrons among the discrete states in the forbidden gap, in the manner described

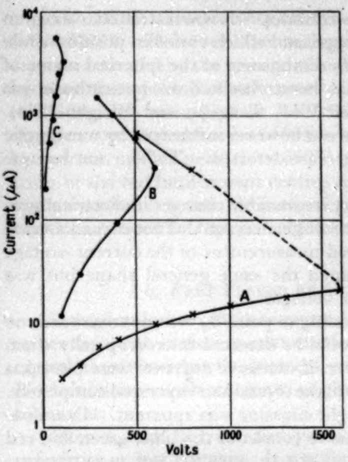


Figure 1. Current–voltage characteristics for a crystal of CdS in the dark. Curve A was measured first; after A had been obtained and the applied voltage removed, curve B was measured. Curve C was measured in a similar way after curve B had been taken.

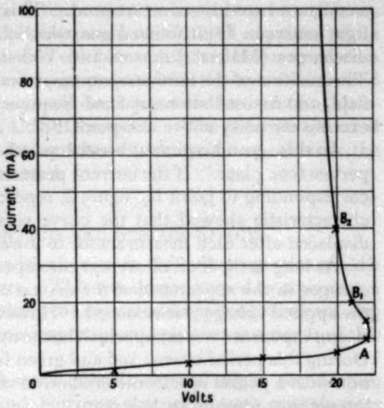


Figure 2. Current–voltage characteristic for a conducting crystal of CdS. At point A the crystal temperature rises; between B_1 and B_2 sustained red light emission is observed; for currents in excess of B_2 the crystal is destroyed and a vapour arc is set up.

by Smith and Rose (1955) and Rose (1955). As a result of the passage of high currents after breakdown a non-equilibrium condition is created in which electrons are trapped in high-lying energy states. In consequence the quasi Fermi level is raised and the conductivity is increased. After the applied voltage is removed the trapped electrons are thermally excited to the conduction band, whereupon thermal equilibrium can be established. This process is slow at room temperature in the dark. Heating or irradiation with light would be expected to speed up the return to thermal equilibrium, and so it proves in practice. The temporary change in conductivity can be eradicated in a few minutes either by heating to 300°c or by irradiation with white light.

§ 4. The Second Breakdown Process

The second stage of breakdown was obtained by applying an increasing direct voltage to a small load resistor and the specimen in series. The specimen had already been subjected to the first breakdown to reduce its dark resistivity to a value of 1–10 ohm cm.

The current–voltage curve in figure 2 is a typical illustration of the behaviour observed. At the point A the crystal temperature began to rise, as evidenced by the change in colour, and at point B_1 sustained light emission was observed. This emission was red in colour and when observed with a pocket spectrometer, was found to consist of a broad band extending from the yellow-green to red, cutting off at about 6500Å. The luminescent emission did not extend throughout the whole region of the crystal between the electrodes. It was restricted to a region which appeared circular under the microscope, and which varied in position while a constant current was maintained. This is reminiscent of the spherical region of light emission from ionized gas which has been observed in hot-cathode gas discharges (Malter, Johnson and Webster 1951, Pengelly and Wright 1954). The position of the luminescent region was not however influenced by a magnetic field, and no oscillations of fixed frequency were detected. The current became excessively noisy above the point B_1.

In this second region of breakdown also, irreversible changes in electrical properties took place. If the current passed through a crystal did not exceed a value corresponding to point B_2, figure 2, repeated measurement of the current–voltage characteristic showed that the curve retained the same general shape but was displaced after each measurement to lower voltages.

As long as the level of current corresponding to point B_2 was not exceeded, no changes in the appearance of the crystal could be detected microscopically when the applied voltage was removed. However, if excessive currents were passed, a strong vapour arc was set up which lasted until the crystal had vaporized completely. During this period intense red and green light emission was apparent. Examination with a pocket spectrometer showed this to consist of the blue, green and red lines of the cadmium arc spectrum.

§ 5. SPECTRAL DISTRIBUTION OF PHOTOCONDUCTIVITY BEFORE AND AFTER BREAKDOWN

Spectral distributions of the photo-current before and after passage of a large current through a typical crystal are shown in figure 3. With an insulating CdS crystal prior to breakdown a large part of the photo-response is confined to a band in the green with its maximum at $\lambda = 5150$Å. The photo-response falls off rapidly as the wavelength of the incident light is increased. After activation by passing a large current, the larger part of the photo-response is still confined to the band in the green, but now there is a considerable response at longer wavelengths extending to the near infra-red. A fairly sharp long-wave threshold of photoconductivity at $\lambda = 8500$Å was observed in all specimens. We have observed that

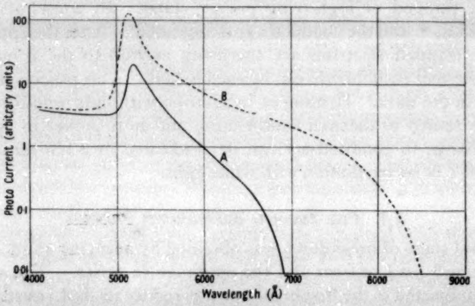


Figure 3. Spectral distribution of photo-current of a CdS crystal before and after breakdown. A, before breakdown, 400 volts applied; B, after breakdown, 20 volts applied.

crystals of CdS activated with chlorine also show a threshold of photoconductivity at $\lambda = 8500$Å. However, the analogy between chlorine activated crystals and crystals activated by the passage of high current cannot be carried too far. This is because the spectral distribution of photocurrent of CdS:Cl crystals, while exhibiting the main peak at $\lambda = 5150$Å also exhibits a smaller, but nevertheless significant, peak at about $\lambda = 6500$Å, which is not present with the crystals activated by current. This peak at $\lambda = 6500$Å may well correspond to the excitation spectrum of the red luminescent centres of CdS:Cl crystals. The chlorine activated specimens have a red luminescence under ultra-violet excitation whereas the crystals activated by high current flow do not.

§ 6. PART PLAYED BY CONTACTS AND SURFACE EFFECTS

Activation by current is associated with changes in the bulk properties of CdS rather than with surface or contact effects. This was demonstrated by activating a crystal, then removing the contacts, etching the crystal and applying new contacts. No reduction in dark resistance or photo-response was apparent as a result of this treatment.

After a high current had passed some crystals showed slight rectification properties at low voltages. It was found, however, that when such crystals were cleaned and new contacts applied, no rectification effects could be observed. The crystal then showed the lower dark resistance typical of a specimen activated by current.

To check that the effects observed were not due to surface contamination brought about by heating in air the whole series of experiments was repeated with crystals in vacuum, in sealed-off gettered tubes. There was no significant difference in the results obtained.

By using fired silver contacts it was possible to bake a tube under vacuum at 400°c. It was found that the rather poor photoconducting properties of CdS crystals as prepared were unaffected by evacuation at room temperature. After baking, however, the photo-response was found to have increased one hundredfold. Nevertheless, greater response still could be obtained by giving the breakdown treatment. Baking a virgin crystal in air at 400°c did not have any activating effect. It seems likely, then, that baking in vacuum must lead to some modification of the crystal surface. It may well be that the passage of a large current through a crystal leads to a similar modification of surface properties.

§ 7. DISCUSSION

The results reported here are substantially in agreement with the ideas developed by Diemer. Thus the initial breakdown of a CdS crystal under strong field conditions is a field breakdown mechanism analogous to ionization by collision in gases. As a result, a large current flows and the temperature rises. This causes a further reduction in resistance, giving rise to a negative slope current–voltage characteristic. During this time irreversible changes in specimen resistance are taking place and the resistance falls so low that a smaller wattage is dissipated in the crystal and its temperature falls. Owing to the change in resistivity, however, the current flow can be maintained with the crystal near room temperature.

The second stage of breakdown is completely thermal in character and is brought about by Joule heating when very large currents pass through the crystal. If this heating becomes excessive a vapour arc is obtained.

Perhaps the most interesting feature of this work is that CdS crystals can be activated to become excellent photoconductors simply by passing large currents. Activation is usually effected chemically (Bube and Thomsen 1955, Kröger, Vink and van den Boomgard 1954) when dark resistivities ranging from 10^{12} to about 10^{-2} ohm cm can be obtained by adding controlled quantities of chlorine or aluminium, for example. To explain these results Kröger *et al.* have developed a model which depends on the introduction of discrete levels into the forbidden gap on the energy band scheme. Some such levels must also be introduced into the crystal during dielectric breakdown.

Measurements of the spectral distribution of photoconductivity of crystals activated by current and of crystals activated with chlorine both show the same long wavelength threshold of photoconductivity at $\lambda = 8500$Å. However, before a model can be developed for the energy levels introduced with current activated crystals, many more properties must be examined. Thus it is hoped that measurements of infra-red quenching effects, which are at present in progress, can be correlated with measurements of the thermally stimulated current, and with measurements of the electrical conductivity and Hall effect over a wide range of temperatures.

References

Boer, K. W., and Kümmel, U., 1952, *Z. Phys. Chem.*, **200**, 193.

Boer, K. W., Kümmel, U., and Rompe, R., 1952, *Z. Phys. Chem.*, **200**, 180.

Bube, R. H., and Thomsen, G. M., 1955, *J. Chem. Phys.*, **23**, 15.

Diemer, G., 1954, *Philips Res. Rep.*, **9**, 109.

Frerichs, R., 1947, *Phys. Rev.*, **72**, 594.

Kröger, F. A., Vink, H. J., and van den Boomgard, J., 1954, *Z. Phys. Chem.*, **203**, 1.

Malter, L., Johnson, E. O., and Webster, W. M., 1951, *R.C.A. Rev.*, **12**, 415.

Pengelly, A. E., and Wright, D. A., 1954, *Brit. J. Appl. Phys.*, **5**, 391.

Rose, A., 1955, *Phys. Rev.*, **97**, 1538.

Smith, R. W., and Rose, A., 1955, *Phys. Rev.*, **97**, 1531.

Brittle Fractures near Equilibrium

By F. C. ROESLER

Imperial Chemical Industries Limited, Akers Research Laboratories, Welwyn, Herts.

MS. received 26th March 1956

Abstract. Conical indentation fractures in silicate glass are stable, so that the speed at which they spread can be controlled. The validity of the Griffith fracture condition (energy balance) has been tested for such fractures and confirmed, with qualifications. Measurements on cone fractures are a method for determining accurately the specific fracture energy of brittle materials under physical and chemical conditions that can be chosen at will.

---

§ 1. Introduction : Unstable and Stable Fractures

When a brittle body breaks its total surface area increases. There is energy associated with the new surface: in an atomic picture of matter it is the energy of the unsaturated bonds of the atoms in the surface layer, which have no neighbours in one direction. This additional surface energy must be balanced by the work of external forces or by a decrease of some form of energy in the system, for instance by a decrease of strain energy.

This argument is basic in Griffith's (1920, 1924) 'theory of rupture'. But it is not the validity of the energy balance which is the central idea, since the first law of thermodynamics is not at issue. The point is in what might be called 'Griffith's hypothesis', that is the assumption that the balance of surface energy and mechanical energy controls the fracture process. This conception, confirmed empirically for the brittle fracture of glass, is striking because in many other fracture processes the surface energy is unimportant.

It is difficult to make a complete list of the conditions under which Griffith's hypothesis is likely to be true. One condition is that the material should be brittle, because when there is plastic deformation the energy that is dissipated is much larger than the surface energy. Another condition is that the fracture must grow slowly, because the energy of the stress waves that are associated with a running crack is also larger than the surface energy. Strictly, it might be argued that the concept of equilibrium between mechanical energy and surface energy demands infinitely slow fracturing, a quasi-reversible process. Such a quasi-equilibrium can in most cases exist for a single moment only, because in most fracture processes the total load is fixed and as the crack develops this load is concentrated on a smaller and smaller cross section. In cases where Griffith's theory applies the stable or unstable behaviour of a fracture is connected with the dependence of the various forms of energy in the system on the crack length. The comparison of the first derivatives of the energies gives the Griffith balance condition. The second derivatives determine whether the crack will, once it is started, spread catastrophically; if with increasing crack length the supply of mechanical energy increases faster than the demand for surface energy then the crack is unstable and runs.

There is a type of brittle fracture that can be produced without intricate straining devices and is stable in the sense that the larger the crack is, the higher is the load necessary to make it grow further. It is the cone fracture of glass named after Hertz (1882). When a spherical indenter is pressed against a flat-surfaced glass specimen it makes the glass break in a ring crack around the contact area. When the pressure grows the ring crack spreads downwards, flares out and tends in shape to the mantle of a circular cone. It is not implausible that this fracture is stable, since as the crack spreads the area of the cone base increases and the load becomes more distributed. Now it is plain that here is a case which should give a good test of Griffith's hypothesis, since the fracture can be grown slowly. If the hypothesis is borne out then experiments on cone fractures are perhaps a good method for measuring the surface tension of brittle materials.

§ 2. EXPERIMENTAL DETAILS AND TIME EFFECT

Spherical indenters do not produce large and regular cone fractures, since the contact area grows with increasing load and this causes secondary fractures. Cylindrical, flat-ended punches are better. For strength the cylindrical part of the punch used was a short stump backed by a larger piece of steel; the tip and the surrounding face were hardened and a hydraulic ram was used for loading. The specimens were blocks of glass 2 in. × 2 in. × $1\frac{1}{2}$ in., polished on the 2 in. × 2 in. faces and with two opposing 2 in. × $1\frac{1}{2}$ in. faces ground flat. The side faces were smeared with liquid paraffin to make them transparent during the experiment. A side view of a specimen with a large crack, similar to the view seen during the measurements, is shown in the Plate. Usually the edge of the crack is clearly visible in ordinary light, but in some cases of small cracks it was useful to observe the specimen between sheets of polaroid, so that the stress concentrations at the crack edge could be seen. The estimated errors of measurement are $\pm 0{\cdot}5\%$ for the projection of the crack diameter s, which was measured with a travelling microscope, and $\pm 1\%$ for the indenter force P, which was read from a high quality pressure gauge. The actual scatter of the results for the quantity $s/P^{2/3}$ was of the order of $\pm 2\%$, which indicates that the specimen properties did not vary much.

Figure 1 shows a cone fracture schematically. In the theory below it is assumed that the specimen is large compared with the crack, so that the problem becomes equivalent to that of a cone crack in an infinite half-space of elastic material. In practice specimens of moderate size apparently give valid results, because the crack remains straight and regular even when it becomes relatively rather large. The fracture shown in the Plate is about the size where the relation of the indenter force to the crack diameter begins to deviate from that for the smaller fractures. When a crack is made to approach the side faces of the specimen quite closely then it tends to curve upwards, in hat-brim fashion.

At a certain stage in the theory it will also be assumed that the fracture diameter s is large compared with the indenter diameter δ. In the experiments only ratios s/δ less than about 7 could be realized, because at a certain limiting load the glass crushes under the indenter tip. (When this happens the indenter may be spoiled. The ranges of safe loads for the indenters and the glass used can be read from figure 4 below.) In what sense values $s/\delta \simeq 7$ can be regarded as large will become clear later, from the observed relation between the force and the crack diameter.

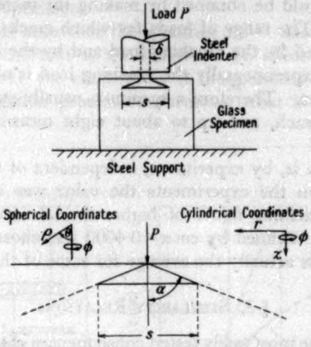


Figure 1. Cone fracture schematically, and as idealized for theoretical treatment.

If the cone fracture obeys Griffith's theory and the surface tension of glass is a fixed quantity, then for any given load the crack should spread to a certain diameter and no further, independent of the time of loading. In reality things will hardly be so simple, even if the theory is true, because the surface tension will change with time. Such changes might still occur if the newly created surface was chemically pure and was kept under perfect vacuum, since there might be processes of diffusion and orientation in the glass itself. In the experiments the crack was open to the atmosphere, so that air and moisture must have entered it. Furthermore, traces of paraffin might have crept into the crack along the surface, although visibly the upper, indented, face of the specimen appeared clean. Figure 2 shows the observed growth of the crack in time, under a constant indenter load. The growth curves found with several specimens are consistent within the accuracy of the figure. Apparently the chemical state of the crack surface was the same in each experiment. After 15 minutes at constant load the crack diameter changes slowly enough to be measured conveniently with the equipment used. For this practical reason 15 minutes was taken as standard loading time; results given below without qualification are 15-minute values.

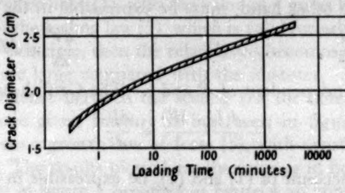


Figure 2. The growth of the crack in time, at constant load.
$P = 1570 \times 10^6$ dyn, $\delta = 3 \cdot 175$ mm.

When a crack is grown stepwise by applying successively higher loads, care must be taken that the crack size attained with each load in 15 minutes is independent of the earlier loading history. Within the accuracy of the experiments this independence could be obtained by making the increments of the load not smaller than 10%. The range of loads for which cracks can be observed with any indenter is limited by the crushing load and by the load at which the first ring crack forms. Experimentally the crushing load is about twice the load at which the crack starts. Therefore a specimen usually stands about seven load increments of 10% each, and up to about eight measurements can be made with each specimen.

The crack angle α is, by experiment, independent of the indenter diameter. For the glass used in the experiments the value was $\alpha = 68{\cdot}5^\circ \pm 1^\circ$. In the computations some definite figure of higher formal accuracy had to be used and for this the value defined by $\cot\alpha = 0{\cdot}4000$ was chosen. This corresponds to $\alpha = 68^\circ\, 12'$, which is actually the average for some of the larger cracks.

§ 3. SIMILARITY RELATIONS

Scaling laws are the most easily tested consequences of the Griffith hypothesis. For a given material the crack angle α is invariant, so that two cone fractures with diameters s, s' and produced by indenters with diameters δ, δ' are geometrically similar if $s/\delta = s'/\delta'$. More generally, for a given specimen material and indenter type, all cone fractures for which s/δ has a constant value are geometrically similar. For such fractures with a given value of s/δ, scaling relations can be written down for any variable connected with the cone fracture by considering the dimension of the variable in question with respect to P and s. For instance, by considering in turn the dependence of the stresses, of the local strain energy and of the volume on P and s, it is found that the total strain energy U must be expressible in the form $U = HP^2/s$, where H is independent of P and s. Again, the total surface energy V must be expressible in the form $V = Ks^2$, with K independent of P and s. The derivatives, with reference to the crack diameter, of the strain energy and of the external work W on the one hand, and of the surface energy on the other hand, must be expressible in the forms

$$\frac{d(U+W)}{ds} = \frac{LP^2}{s^2}, \qquad \ldots\ldots(1) \qquad\qquad \frac{dV}{ds} = Ms, \qquad \ldots\ldots(2)$$

where L and M are independent of P and s. The Griffith balance condition is

$$-\frac{d(U+W)}{ds} = \frac{dV}{ds} \qquad \ldots\ldots(3)$$

and it must, on the account of (1) and (2), be expressible in the form

$$P^2/s^3 = -M/L = N \qquad \ldots\ldots(4)$$

or

$$s/P^{2/3} = N^{-1/3} = n \qquad \ldots\ldots(5)$$

where N and n are independent of P and s. The scaling law (5) can be tested by establishing the empirical relation between $s/P^{2/3}$ and s/δ for a number of indenters of different size. If the facts agree with (5) the same empirical function should be found for all indenters. Figure 3 shows the result, for three indenters with $\delta = \frac{1}{4}$ inch, $\delta = \frac{1}{8}$ inch and $\delta = \frac{1}{16}$ inch. Eleven specimens were used and this gave about 15 points for each indenter size. The scaling law (5) appears to be valid. In the original large-scale plot with individual points, the results due to indenters of different size superpose as well as those for one indenter of fixed size and a number of specimens.

From figure 3 it is seen that the quantity $s/P^{2/3}$ does not vary much and appears to tend to a constant as s/δ increases. In effect when s/δ is large the force necessary to make the crack grow is independent of δ. This indicates that the crack propagation process conforms to what is known in the theory of elasticity as St. Venant's principle. Now the limit $s/\delta \to \infty$ corresponds to an idealized fracture where the core ends in a sharp point (the vertex of the cone) to which the indenter force is applied. The idealized fracture is geometrically self-similar, when it is considered in different stages of growth, and these stages should therefore obey (5), with the value of n given by the limit in figure 3. In other words: if the scaling law (5), which is valid for arbitrary s/δ, is combined with St. Venant's principle, then the relation (5) becomes the law for the growth of fractures that are large compared with the indenter.

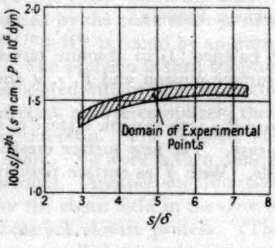


Figure 3. Test of scaling law (5). The experimental points for indenters of varying size fall together.

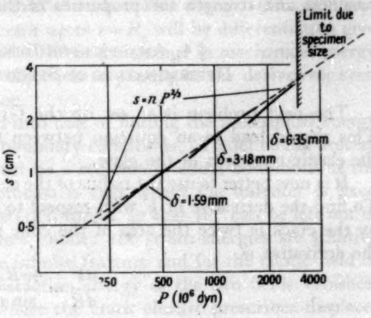


Figure 4. Crack diameter as function of the indenter load (15-minute values).

The correspondence between the scaling law for different fractures and the growth law for one given fracture is best seen in figure 4. In this double logarithmic diagram a growth law of form (5), with constant n, gives a straight line of slope 2/3. The results plotted are the same as in figure 3. The approach to a constant level in figure 3 becomes in figure 4 an approach of the empirical growth curves to the asymptotic growth law of the idealized fracture. The ranges of fracture diameters for the individual indenters and the unreduced values of s and P can now be seen.

The level of the asymptote which gives the best fit in figure 4 can be determined by going back to the more accurate plot in figure 3. The reduced variable $100s/P^{2/3}$ attains a limiting value $1{\cdot}57 \pm 0{\cdot}03$. It follows that in figure 4 the asymptote should cut the line $s = 2$ cm at a load $P = 1438 \times 10^6$ dyn. This pair of quantities will be used to find the surface tension.

In the concept of the idealized cone fracture, where the indenter is replaced by a singular force, all references to the properties of the indenter are omitted. If the size of the indenter tip becomes unimportant then the same should be true for the shape of the tip and the properties of the indenter material. The theory of the idealized cone fracture is therefore only concerned with the elastic field in the glass specimen. In so far as fractures with s/δ much greater than 1 can be made and the theory is true for them, such fractures must be governed by the properties of the glass only and by the chemical condition of the crack. This is a useful conclusion, since in many other indentation type measurements of hardness and strength the properties of the indenting tool are very important.

§ 4. ANALYSIS FOR IDEALIZED FRACTURE. DETERMINATION OF SPECIFIC SURFACE ENERGY

The next problem is to set up the Griffith balance (3) in absolute terms. This should lead to an equation between the surface tension and P, s, α and the elastic constants of the glass.

It is now better to use the radius of the cone base $R=\frac{1}{2}s$ as variable instead of s. To find the derivative of V with respect to R is easy. The new surface created by the crack is twice the area of the cone mantle. With T as surface tension, the derivative is

$$\frac{dV}{dR}=\frac{4\pi RT}{\sin\alpha}. \qquad \ldots\ldots(6)$$

The derivative of the mechanical energy $d(U+W)/dR$ is more difficult to find. Since the exact solution for the elastic field round the cone crack is not known, an approximation method must be used; but before the problem can be treated by approximation it must be transformed. Some discussion is necessary to justify and explain this, and the argument is a little involved. The approximation which is used eventually is a well-understood and systematic method (Rayleigh–Ritz), and in principle it is possible to find $d(U+W)/dR$ with arbitrary accuracy. The computation is, however, long, even for moderate accuracy, and it has therefore only been done for the one value of α observed in the present experiments, and for the one value $\nu=\frac{1}{4}$ of Poisson's ratio. (This is a mean value for silicate glass. The exact value of ν for the glass used in the experiments was not known.) The final result for the Griffith balance is given in equation (10) below.

To try a direct approximation method for the elastic field would be futile as far as the determination of $d(U+W)/dR$ is concerned. The change in the mechanical energy of the system that is caused by the cone crack is only a small fraction of the total mechanical energy, so that $d(U+W)/dR$ is small compared with $(U+W)/R$. If an approximation to the total mechanical energy were set up and $d(U+W)/dR$ determined from it by differentiation, the result might be quite wrong even if the approximation to $U+W$ itself were relatively good.

This difficulty can be avoided by basing the approximation on a comparison. The crack field can be compared with a known, exact solution, which is due to Michell (1900). It describes the elastic field in an infinite circular cone which is loaded by a singular force at the vertex. If an idealized cone fracture, as shown in the lower part of figure 1, were to extend to infinity, then the core of this fracture would be a cone in the state worked out by Michell, and the surrounding rim of material would be free of stress. The shape of the gap or crack between the rim and the core is, for this infinite fracture, given by the displacements of the mantle of Michell's cone, since the rim has vanishing displacements. On the other hand for a finite cone crack of radius R the relative displacements of the mantle elements of rim and core must vanish for r greater than R, by the definition of R. The generation of a finite cone crack can then be decomposed in imagination: first the crack extends to infinity, then it is closed again from infinity to $r=R$; the displacements, at the mantle surfaces, which are necessary for the closure, are known from Michell's work. The energy, U^*+W^* say, which is associated with closing the crack up to $r=R$, will by differentiation give $d(U+W)/dR$. Since the whole of U^*+W^* is the change in mechanical energy caused by the crack there is now no objection to passing to the derivative, even if U^*+W^* is found by approximation.

The problem is then transformed into that of finding the mechanical energy associated with certain displacement boundary conditions. To define the process of crack closure completely there must be a stipulation about conditions at the indenter tip. The simplest ideal process is that where the indenter is fixed rigidly. No work is then done by external forces, so that W^* can be left out of account. Another simplification comes about: the strain energies are additive for the elastic field in the core of the infinite fracture and for the elastic field of the crack closure process. (The interaction energy of the two fields vanishes, because all those surface elements where the crack closure prescribes displacements are along the mantle, and therefore they are by definition free of stress in the state of the infinite fracture.) For this reason the energy U^* associated with the crack closure is defined completely by the prescribed boundary displacements. For a physical picture of this result, it might be imagined that the core of the infinite fracture is modelled in stress-free glass. The process of crack closure can be carried out by fitting this initially stress-free replica into the rim and then closing the gap along the mantle surfaces for r greater than R. At this stage the condition that the vertex of the core shall be held fixed becomes inessential and can be dropped, since keeping the condition would not increase U^* by a finite amount.

From the self-similarity of the idealized fracture U^* must be inversely proportional to R. The dependence of U^* on P and on the modulus of the glass is also obvious, and in effect U^* can be written

$$U^*=\omega(\alpha,\nu)P^2/GR. \qquad \ldots\ldots(7)$$

G is the shear modulus and ν is Poisson's ratio. $\omega(\alpha,\nu)$ is a number which gives the value of U^* for $P=R=G=1$, and this number must be found by an approximation method. A summary description of the procedure by which ω was computed, using the Rayleigh–Ritz method, is given in the Appendix. The result of the numerical work is

$$\omega=7{\cdot}45\times10^{-3} \qquad \ldots\ldots(8)$$

for $\alpha=68^\circ\,12'$ and $\nu=\frac{1}{4}$. From the approximation method used, (8) is an upper bound, which means the true value is not higher. There is no precise information about the error, but from the trend of the computation it appears unlikely that the error exceeds a few per cent and it may be less than one per cent. Of course, no final and absolute confidence can be placed on the value of ω before it has been confirmed by an entirely independent computation.

From (7) follows the derivative

$$\frac{dU^*}{dR} = -\omega\frac{P^2}{GR^2}, \qquad \ldots\ldots(9)$$

and from the argument by which U^* was introduced, the derivative dU^*/dR is equal to the derivative of the mechanical energy $d(U+W)/dR$. The Griffith balance can then be written, by combining (6), (9) and (8),

$$\frac{dV}{dR} = -\frac{d(U+W)}{dR} = -\frac{dU^*}{dR}, \qquad \ldots\ldots(10)$$

or explicitly

$$\frac{4\pi RT}{\sin\alpha} = \omega\frac{P^2}{GR^2}. \qquad \ldots\ldots(11)$$

This can be solved for T, and with the numerical value of ω put in becomes

$$T = \frac{7{\cdot}45\times10^{-3}P^2\sin\alpha}{4\pi GR^3}. \qquad \ldots\ldots(12)$$

With $P = 1438\times10^6$ dyn corresponding to $R = 1$ cm (experimental values), $\sin\alpha = 0{\cdot}9285$, and $G = 2{\cdot}8\times10^{11}\,\mathrm{dyn\,cm^{-2}}$ (a mean table value for silicate glass), the surface tension becomes

$$T = 4065\ \mathrm{dyn\,cm^{-1}} \simeq 4100\ \mathrm{dyn\,cm^{-1}}, \qquad \ldots\ldots(13)$$

since keeping more than two figures would not be reasonable. From the data that were used, (13) is the '15-minute value'. Since the third power of R in (11), the growth in time of the crack, as shown in figure 2, implies strong changes in the surface tension. Twenty seconds after loading the value of T is twice the '15-minute value'. After the load has been on for several days the value of T is half the '15-minute value'.

§ 5. DISCUSSION

If the structure of a solid is known, the surface tension can be estimated from the number and strength of the chemical bonds which must be broken to create a new surface. Such a calculation gives the total surface energy at the absolute zero of temperature, and it gives an over-estimate, since there will be rearrangement in the surface layer with partial saturation of the ruptured bonds. To find the surface tension at finite temperatures (the free surface energy in the sense of thermodynamics) requires a further correction. The results of some such estimates for valence-bond crystals are given in the Landolt–Börnstein tables (6th edition, Volume I/4, p. 543). The method can hardly be applied to the network structure of glass, since here the rearrangement in the surface layer may be decisive for the value of T. At best the count of broken bonds gives in the case of glass a crude upper limit, perhaps $T < 3000\,\mathrm{dyn\,cm^{-1}}$ for a clean surface. The value (13) of T derived from the cone crack experiments therefore seems a little large, but it is of the expected order of magnitude.

The value (13) can also be compared with the results of Griffith's own experiments on cracked glass tubes burst by hydraulic pressure. Griffith had measured the surface tension of his glass in the liquid state, and extrapolated the results to room temperature. In this way he had found a value $T = 546$ dyn cm^{-1}, and he investigated whether the bursting experiments might give the same value. This was true when the annealing of the cracked tubes was carried out at a certain temperature and for a certain time, but the result for T depended rather critically on the annealing procedure (Griffith 1924). (It can be seen from Griffith's figures that very thoroughly annealed tubes gave values for T up to eight times the value found with unannealed tubes. The aim of the annealing procedure was to remove the stress due to scratching but without softening the glass so much that the corners of the crack enlarge. Various methods of annealing were tried, and compared with respect to the consistency of the results.) One procedure tried by Griffith, which involved more thorough annealing, gave well reproducible results from which a value $T = 1800$ dyn cm^{-1} can be deduced. This agrees, approximately, with the values of T found by the cone crack method for very long loading times.

The time effect observed in the cone crack experiments is an indication that the fracture process involves some losses by elastic hysteresis or viscosity. From figure 2, the value (13) and equation (12) it follows that T is of the order 10 000 dyn cm^{-1} for loading times of a few seconds. This seems too high for a true surface tension, even if the surface were absolutely clean. It is a fact that glass shows considerable elastic hysteresis at room temperature, for stress cycles of a few minutes' duration. As the edge of the growing crack transfers the stress field through the material, deviations from perfect elasticity would cause a loss of mechanical energy. There is direct evidence for the non-ideal behaviour of the glass in the fact that the crack remains open by a few microns after the indenter load has been removed. From all these arguments it appears likely that part of the time effect shown in figure 2 is due to anelastic processes, and only the remainder is due to the lowering of the surface tension by adsorption of vapours and other chemical changes.

Incidentally, the knowledge of the existing anelastic effects makes it uncertain what values of the elastic constants should be used to improve on the crude assumptions ($\nu = 0{\cdot}25$, $G = 2{\cdot}8 \times 10^{11}$) made above. Accurate values for the elastic constants of glass are usually determined at ultrasonic frequencies by optical methods, and such values may not apply for stress cycles of the order of hours.

Despite these uncertainties it seems that cone crack experiments are an attractive testing method for glass. Chemical influences on T can be studied by putting liquid round the indenter tip, so that the atmosphere pushes it into the opening crack. Some such experiments have been made and it was obvious that there are strong effects. For instance if hydrofluoric acid is put into the crack then even small loads give large cracks in a short time. The effect is not quite trivial, since ordinary etching action should not occur preferentially at the crack edge. Of course, for comparative measurements of T on the same glass it is not necessary to know the value of ω.

The method may be extended to other materials that are sufficiently brittle or can be made so by the right conditions of temperature or rate of loading. Mr. D. G. Christie of these laboratories has been able to produce fairly clean cone cracks in 'Perspex' (a form of polymethylmethacrylate) by using, at room temperature, a high rate of loading. It must be admitted that the less brittle the material is, the more would the apparent value of T be due to dissipative processes. Yet the property that is measured by the cone crack would still be independent of the indenting tool and would be characteristic of a crack propagating in clean, virgin material.

ACKNOWLEDGMENT

I wish to record my thanks to Dr. H. Kolsky for many discussions on the subject of this paper.

REFERENCES

GRIFFITH, A. A., 1920, *Phil. Trans. Roy. Soc.*, **221**, 163; 1924, *Proc. Int. Congr. Appl. Mech.* (Delft), 55.

HERTZ, H., 1882, *Verh. Ver. Gew. Fleiss. Berl.*, **61**, 449; reprinted 1895 in *Collected Works*, **1**, 174 (Leipzig: Barth).

MICHELL, J. H., 1900, *Proc. Lond. Math. Soc.*, **32**, 23.

PRAGER, W., and SYNGE, L. J., 1947, *Quart. Appl. Math.*, **5**, 241.

APPENDIX

The evaluation of ω is a minimization problem. Among all continuous displacement fields which satisfy a given set of displacement boundary conditions the true field is the one which makes the strain energy a minimum. The proof can be found in various textbooks on elasticity. (One demonstration which I found particularly helpful occurs in Prager and Synge (1947): their paper will also supply the precise meaning of the term 'set of displacement boundary conditions'.)

Michell worked with spherical polar coordinates; figure 1 shows the notation used here. In the special case $\nu=\frac{1}{4}$ and for $P=G=1$ his formulae give for the displacements u_θ, u_ρ at the cone mantle

$$2\pi\rho u_\theta=\frac{-(3/2)\sin\alpha}{(2+\cos\alpha+2\cos^2\alpha)(1-\cos\alpha)} \qquad \ldots\ldots(14)$$

$$2\pi\rho u_\rho=\frac{3\cos\alpha-\cos^2\frac{1}{2}\alpha}{(2+\cos\alpha+2\cos^2\alpha)(1-\cos\alpha)}. \qquad \ldots\ldots(15)$$

These displacements define the width of the crack that is to be closed for $1<r<\infty$. There are no other prescribed displacements, except that for $\rho\to\infty$ all displacements must vanish.

The expressions for the strains are simpler in cylindrical than in polar coordinates, and therefore it is easier to work out the approximation in cylindrical coordinates.

The displacements are transformed by

$$u=u_r=u_\rho\sin\alpha+u_\theta\cos\alpha=\frac{M}{r}, \qquad \ldots\ldots(16)$$

$$w=u_z=u_\rho\cos\alpha-u_\theta\sin\alpha=\frac{N}{r}. \qquad \ldots\ldots(17)$$

The quantities M and N are to be taken as defined by (16) and (17). M and N are independent of r. In working out the values of M and N it must of course be remembered that $r = \rho \sin \alpha$.

For $u_\phi = 0$, from symmetry, the strains in cylindrical coordinates are

$$\left.\begin{aligned} \epsilon_r = \frac{\partial u}{\partial r}, \qquad \epsilon_\phi = \frac{u}{r}, \qquad \epsilon_z = \frac{\partial w}{\partial z}, \\ \gamma_{rz} = \frac{\partial u}{\partial z} + \frac{\partial w}{\partial r}, \\ \epsilon = \epsilon_r + \epsilon_\phi + \epsilon_z. \end{aligned}\right\} \qquad \ldots\ldots(18)$$

For $G = 1$ the local strain energy becomes

$$A = \epsilon_r^2 + \epsilon_\phi^2 + \epsilon_z^2 + \tfrac{1}{2}\gamma_{rz}^2 + \left(\frac{\nu}{1-2\nu}\right)\epsilon^2. \qquad \ldots\ldots(19)$$

The total strain energy is the integral of A over the volume of the body.

To obtain workable expressions the field must be built up piece-wise, dividing the whole body into a number of domains. All displacement functions must either be matched at the inter-domain boundaries or be chosen so that they vanish there. Four domains were taken, as follows (β is written for $\cot \alpha$):

domain I: $r<1$, $z<\beta r$ ('near rim'),
domain II: $r>1$, $z<\beta r$ ('far rim'),
domain III: $r<z/\beta$, $z>\beta$ ('far core'),
domain IV: $r<z/\beta$, $z<\beta$ ('near core').

Now it is to be remembered that the prescribed displacements (16) and (17) are relative displacements of the two mantle surfaces, by the physical meaning of the crack closure process. Therefore they are the differences of the absolute displacements in domain II and domain III, at the mutual boundary of these domains. At the boundary II/III the displacements then must satisfy

$$u_{\mathrm{II}} - u_{\mathrm{III}} = \frac{M}{r}. \quad \ldots\ldots(20) \qquad\qquad w_{\mathrm{II}} - w_{\mathrm{III}} = \frac{N}{r}. \quad \ldots\ldots(21)$$

Displacement fields which fit these conditions can in domain II be built up by using for u and w components of type $1/r$, z/r^2, z^2/r^3, ...; and in domain III they can be built from components of type $1/z$, r/z^2, r^2/z^3, ... The choice of components for domain II can be refined a little by taking account of the stress boundary conditions which the true field must satisfy for $z=0$. Again, when making the choice for domain III some components can be rejected because they would involve strain singularities at $r=0$.

All components which survive must be matched at the boundaries II/I and III/IV by suitable continuations in domains I and IV. It is already becoming evident that the total number of functions is inconveniently large, but more components must yet be introduced, to give the field sufficient freedom near the crack edge and in the 'near rim' and 'near core' domains. These additional fields must be so designed that they give zero displacements along the boundary II/III, and they must vanish with increasing r and z more strongly than the $1/r$ and $1/z$ type components considered earlier.

When the choice of all field components has been made, the strains (18) must be found for each component. The next step is to find for all fields and combinations of fields the self-energies and interaction energies, by forming the local strain energy according to (19), or the analogous interaction energy with products of strains, and then integrating over the relevant domain or domains. Eventually a table of all self-energies and interaction energies is made up. This, together with the boundary conditions given by the quantities M and N, is used to set up equations for the amplitudes of the various components. These equations are the conditions for the total strain energy to be a minimum, and they are linear in the unknown amplitudes.

The computation that was done involved 17 independent components of the displacement field. Since there are two boundary conditions, which can be accommodated by using Lagrangian multipliers, the total number of equations was 19. These happened to form three independent sets of 11, 4 and 4 simultaneous equations, respectively. The three sets were solved, partially by elimination and partially by iteration. Finally, the total strain energy was evaluated from the amplitudes of the minimum.

The result was

$$\omega = 7{\cdot}45 \times 10^{-3}. \qquad \ldots\ldots(22)$$

In an earlier computation with only six independent field components the result was

$$\omega = 8{\cdot}3 \times 10^{-3}. \qquad \ldots\ldots(23)$$

As would be expected, the result of the more elaborate computation is lower and therefore nearer to the true value. Besides giving a check, the comparison of the two results also gives an indication that the error of the lower figure is probably small. If the true minimum were much lower than (22), then the increase in the number of components from 6 to 17 should have changed ω by more than 10%.

The Anisotropy Correction in Ferromagnetic Resonance

By K. J. STANDLEY and K. W. H. STEVENS

Department of Physics, University of Nottingham

Communicated by L. F. Bates; MS. received 10th May 1956

Abstract. A method is described for finding the net absorption, at a given magnetic field, resulting from the superposition of absorption lines from the randomly distributed crystallites of a polycrystalline material. It is then shown that, due to magnetocrystalline anisotropy, the maximum in a ferromagnetic resonance curve does not coincide with the centre of gravity of the absorption line. This is important because the usual resonance equation $h\nu=g\beta H$ implies the use of the centre of gravity field, so that to obtain the true g a correction must be made to the field for maximum absorption. The magnitude of the correction depends on K_1/M and is estimated numerically.

---

§ 1. Introduction

The resonance condition in ferromagnetic resonance is usually expressed in the form $h\nu=g\beta H_{\mathrm{eff}}$, where $h\nu$ is measured directly and where H_{eff} is not the actual external magnetic field H_{m} for maximum absorption, but is an effective field which is related theoretically to H_{m}. Thus g can be regarded as defined by this equation, its value depending on the corrections made to H_{m}. The reason for using H_{eff} is that it is generally assumed that, provided it is chosen correctly, the value obtained for g is characteristic of the ferromagnetic material under investigation, and should not vary with the experimental conditions, such as the measuring frequency, the temperature, the shape of the specimen, etc. The theoretical problem of relating H_{m} and H_{eff} has been studied by many authors (Kittel 1948, Van Vleck 1950) and it is well established that a most important correction arises from demagnetization effects. That this is not enough was pointed out by Okamura, Torizuka and Kojima (1952) who found that, with certain ferrites, a constant value of g was only obtained if an additional internal field H_{A} was added to the H_{eff} obtained with the demagnetizing correction. They found H_{A} to be constant with measuring frequency but variable with temperature. Miles (1954) pointed out that H_{A} might arise, wholly or in part, from magnetocrystalline anisotropy, and while this is almost certainly true his analysis cannot be regarded as entirely satisfactory. Another treatment of the anisotropy correction has been given by Bagguley (1955), for a very special case. From the work of these two authors there seems little doubt that there should be an anisotropy correction in H_{eff}, but as it was not clear just how important it could be we have made a more detailed theoretical study of it.

§ 2. The Theoretical Model

For simplicity we shall assume that our resonance curve is obtained with a polycrystalline sphere of a ferromagnetic material, the individual crystallites of which have cubic crystal structure. The demagnetizing correction then becomes zero and each crystallite gives its own contribution to the full resonance curve. The resonance condition for a crystallite oriented so that the external magnetic field, H_0, direction is (l, m, n) referred to its cubic axes is

$$h\nu = g\beta\left\{H_0 + \frac{5K_1}{M}\left(l^4 + m^4 + n^4 - \tfrac{3}{5}\right)\right\} \qquad \ldots\ldots(1)$$

(Van Vleck 1950) where K_1 denotes the first order anisotropy constant (the others being neglected) and M the saturation magnetization. The observed resonance line consists of the superposition of the lines of all the separate, randomly oriented crystallites which make up the sphere. As Van Vleck pointed out, the position of the centre of gravity of the line so obtained does not depend on K_1, but experimentally it is more convenient to find the position of the maximum, and they do not necessarily coincide. It might be thought that the difference should be easily recognizable, since the implication is that the line is asymmetrical in shape. However, with wide lines and long tails it is not at all easy, experimentally, to find the position of the centre of gravity. The theoretical problem is thus to find the position of the maximum of the absorption curve obtained by superposition, preferably allowing for the fact that the line from each crystallite may also have a finite width. (Bagguley assumed zero width, a much simpler problem.)

We therefore assume that each such line is gaussian in shape, of the form $\exp(-\theta^2/T^2)$, where θ measures the distance in field from H_0, given by equation (1), and T measures its width. This is not the most general assumption which could have been made, for example the width might also vary with (l, m, n), but it is representative, and the method which we develop is capable of handling more complicated shapes, should the necessity arise. We also assume that the whole resonance curve is observed by varying the external magnetic field at constant frequency.

The total intensity of the resonance curve is given by:

$$\int \exp(-\theta^2/T^2)\, d\theta\, dl\, dm\, dn$$

where θ is to be integrated from $-\infty$ to ∞ and the range of variation of l, m and n is such that $l^2 + m^2 + n^2 = 1$. If we now change the variables in this integral in such a way that one of the new ones is H, given by

$$H = \theta + \frac{h\nu}{g\beta} - \frac{5K_1}{M}\left(l^4 + m^4 + n^4 - \tfrac{3}{5}\right),$$

the integration over the remaining variables gives the intensity at the value H of the external field. That is, it performs the superposition of the contributions of the individual resonance lines at the field H. The other new variables, p, ϕ and R are conveniently defined by $p = n$, $\tan\frac{1}{2}\phi = l/m$, $R^2 = l^2 + m^2 + n^2$, in which case the volume element $dl\, dm\, dn$ becomes

$$\tfrac{1}{2}\, dH\, dp\, d\phi\, dR\, \delta(1 - R^2)$$

where the delta function is introduced to allow for the fact that $l^2 + m^2 + n^2 = R^2 = 1$. The intensity at the field H is then proportional to:

$$\int_0^{\pi/2}\int_0^1 \exp\left\{-\left(\frac{1}{g\beta T}\right)^2\left[g\beta H - h\nu + \frac{5g\beta K_1}{M}\left\{(1-p^2)^2\frac{(1+\cos^2\phi)}{2} + p^4 - \frac{3}{5}\right\}\right]^2\right\} dp\, d\phi.$$

The centre of gravity of the line is at $h\nu/g\beta$, and measuring the magnetic field from this position, by setting $g\beta H - h\nu = g\beta x$ we have

$$I(x) = \iint \exp\left\{-\left(\frac{1}{T^2}\right)\left[x + a\left\{(1-p^2)^2\frac{(1+\cos^2\phi)}{2} + p^4 - \frac{3}{5}\right\}\right]^2\right\} dp\,d\phi$$

where $a = 5K_1/M$.

To reduce the computational labour we decided to evaluate this integral for small values of x only, using approximation methods, and found it convenient to set $x = a(0{\cdot}1 - \eta)$, $a/T = \lambda$.

Setting
$$(1-p^2)^2\frac{(1+\cos^2\phi)}{2} + p^4 - 0{\cdot}5 = f(\phi)$$

we have
$$I(0{\cdot}1a - \eta a) = \exp(-\lambda^2\eta^2)\iint \exp\{-\lambda^2[f^2 + 2\eta f]\}\, dp\, d\phi$$
$$= \exp(-\lambda^2\eta^2)\{A + B\lambda^2\eta + C(\lambda^2\eta)^2 + \dots\}.$$

Subject to certain assumptions about the rate of convergence of this series, which we have satisfied ourselves, by numerical estimates, are valid, the maximum absorption occurs near $\eta = B/2A$. Some results of our calculations are given in the following table, where it may be noted that the values in the last column ($xM/5K_1$) are for the position of the maximum measured from the centre of gravity of the curve, and that the values for $\lambda = 0$ and ∞ (Bagguley's case) have been obtained by more elementary methods.

$\lvert\lambda\rvert$	A	B	C	$B/2A$	$xM/5K_1$
0	—	—	—	—	0
1	8·654	1·530	—	0·088	0·012
2	7·861	1·145	—	0·073	0·027
4	6·224	0·406	−0·17	0·033	0·067
∞	—	—	—	—	0·100

$$\lvert\lambda\rvert = \lvert 5K_1/MT\rvert.$$

To appreciate the significance of these results it is instructive to compare iron and nickel, for which K_1/M is $+240$ and -120 Oe respectively (Bozorth 1951) There is some uncertainty about the values to be taken for T, but it would seem reasonable to take it as not very different from 300 Oe for both materials. Then $\lvert\lambda\rvert$ is 4 for iron and 2 for nickel, so that the difference in field between the centre of gravity, for which $h\nu = g\beta H_0$ defines the true g, and the maximum in the absorption curve is -80 Oe for iron and $+16$ Oe for nickel. Thus for iron $h\nu = g\beta(H_{\mathrm{m}} - 80)$, while for nickel $h\nu = g\beta(H_{\mathrm{m}} + 16)$. With measuring fields of order several thousand oersteds and overall line widths some hundreds of oersteds it is clear that the shift in nickel will probably be unobservable. This conclusion is supported by the experimental observation that g for nickel is 2·22 for measurements made at both 3·14 and 1·25 cm wavelength, this value being calculated by assuming that the maximum and the centre of gravity coincide. For iron the g values calculated in the same way are 2·00 and 2·05, a discrepancy which can be removed by including an anisotropy shift of the same order and sign as that which we have estimated, leading to a true g-value of approximately 2·07.

Our conclusion is thus that the anisotropy shift can be of importance, particularly in those substances for which the single crystal line width is small and K_1/M large. Apart from iron, this situation is found in some ferrites, and in a subsequent communication one of us (K. J. S) will present experimental results which demonstrate a shift such as Okamura and his co-workers found. Whether or not the whole of the variation of the uncorrected g's with frequency is due to this effect is not certain, but as a result of this analysis we feel that the anisotropy shift could be the most important.

REFERENCES

BAGGULEY, D. M. S., 1955, *Proc. Roy. Soc.* A, **228,** 549.
BOZORTH, R. M., 1951, *Ferromagnetism* (New York : Van Nostrand).
KITTEL, C., 1948, *Phys. Rev.*, **73,** 155.
MILES, P. A., 1954, *Nature, Lond.*, **174,** 177.
OKAMURA, T., TORIZUKA, Y., and KOJIMA, Y., 1952, *Phys. Rev.*, **88,** 1425.
VAN VLECK, J. H., 1950, *Phys. Rev.*, **78,** 266.

The Magnetic Susceptibilities of Palladium–Uranium Alloys

BY L. F. BATES AND S. J. LEACH[†]
Department of Physics, University of Nottingham

MS. received 5th April 1956

Abstract. Measurements are made of the paramagnetic susceptibilities of palladium–uranium alloys covering the range of solid solution, of an intermetallic compound Pd_3U and of specimens of very pure palladium and uranium over the temperature range 77° to 293°K. Increasing amounts of uranium cause a rapid decrease of susceptibility up to a concentration of 9 atomic per cent. Further increase in the amount of uranium causes a rapid rise in susceptibility, alloys of concentrations greater than 12 atomic per cent obeying a Curie–Weiss law with an effective atomic moment of 3·14 Bohr magnetons. The results are discussed in relation to the properties of the palladium–silver system and close similarity is found for low concentrations assuming that uranium is six times more effective than silver. The temperature-dependent susceptibilities of the alloys of higher uranium content are explained by a change in the valency of uranium leaving two unpaired electrons localized on uranium atoms. An annealed alloy of concentration 17·6 atomic per cent uranium showed an unexplained time dependence of susceptibility.

§ 1. INTRODUCTION

It is not easy to explain the physical properties of an alloy, and in any attempt to do so a considerable number of assumptions has to be made. It is therefore important to measure systematically many properties of alloy systems and then attempt to relate these properties to the theories of the metallic state. The present work is an investigation of the paramagnetic susceptibilities of palladium–uranium alloys through the range of solid solution (0 to 21 atomic per cent), an intermetallic compound Pd_3U and specimens of very pure palladium and uranium. This work is part of a comprehensive study of the properties of palladium–uranium alloys co-ordinated by Dr. P. C. L. Pfeil and Miss J. R. Murray of the Atomic Energy Research Establishment, Harwell.

§ 2. EXPERIMENTAL DETAILS

Temperature dependence of susceptibility from 77°K to room temperature was measured with an apparatus (Leach 1955) developed from that described by Bates, Baker and Meakin (1940), and Bates and Mallard (1950), a vacuum system and a cryostat which enabled accurate measurements of susceptibility to be made from the temperature of liquid nitrogen to that of boiling water being incorporated. The susceptibility standard used to calibrate the apparatus was an aqueous solution of nickel chloride (Nettleton and Sugden 1939). The susceptibility was measured using several field strengths up to a maximum value of 4500 oersteds in order to test for the presence of ferromagnetic impurities. The susceptibility of each specimen was found to be independent of the applied field with the exception of that of the uranium specimen at low temperatures. Special arrangements were made to ensure that the temperature of the specimen could be maintained constant to ± 1 centigrade degree at a chosen value which was uniform along its whole length. The temperature was recorded by a copper–constantan thermojunction placed 1 mm from the lower end of the specimen tube and standardized against three junctions placed inside the tube in a separate experiment.

[†] Now at Safety in Mines Research Establishment, Sheffield.

The palladium specimen, prepared by Messrs Johnson, Matthey and Co. Ltd., was spectroscopically standardized material with the following detected impurities: platinum 0·01%, rhodium 0·002%, osmium 0·001%, iron 0·005%, silver 0·0002%, copper 0·0002%. The uranium specimen contained a total of less than 100 parts per million of metallic impurities. The alloys were prepared from palladium of purity greater than 99·9% and uranium of purity greater than 99·8%. The concentrations of the alloys are listed in table 1. The alloys were made homogeneous by heating *in vacuo* for six hours at 900°C, followed by quenching to room temperature; to eliminate the (unlikely) possibility of the presence of dissolved hydrogen remaining after this treatment and to reduce strain, specimens 1–8 (cf. table 1) were heated for twelve hours at 750°C and then slowly cooled to room temperature. The remaining specimens, with the exception of specimen 11, were not given the latter heat treatment, for the reason indicated in the discussion of the behaviour of specimen 11. The specimens were turned to a size suitable for the measurements, and surface contamination was removed by treatment with boiling (50%) hydrochloric acid.

§ 3. EXPERIMENTAL RESULTS

The mass susceptibilities of the specimens at 90°K and 293°K are given in table 1.

Table 1

I	II	III	IV	V	VI	VII
1(Pd)	0	7·36	5·18	0	—	—
2	0·51	6·03	4·55	3·1	—	—
3	0·92	4·92	3·96	5·5	—	—
4	2·42	2·97	2·71	14·5	—	—
5	6·37	1·01	0·96	38·2	—	—
6	7·25	0·68	0·68	43·5	—	—
7	9·23	0·76	0·71	55·4	—	—
8	12·08	2·05	1·40	72·6	2·87	−417
9	11·1	1·74	1·27	66·6	—	—
10	15·2	5·42	3·15	—	3·19	−221
11	17·6	—	3·71	—	3·09	−168
12	18·5	7·20	4·05	—	3·20	−160
13	†20	9·38	4·60	—	3·12	−115
14(UPd_3)	25·3	13·1	5·83	—	3·09	− 71
15(U)	99·99	1·63	1·70	—	—	—

† Nominal concentration

I, specimen number; II, uranium concentration (atomic %, as analysed); III, measured mass susceptibility at 90°K (e.m.u. $g^{-1} \times 10^6$); IV, measured mass susceptibility at 293°K (e.m.u. $g^{-1} \times 10^6$); V, electron concentration; VI, effective atomic moment (Bohr magnetons); VII, paramagnetic Curie temperature (°K).

The temperature dependence of susceptibility of palladium specimen 1 from 77° to 293°K ,and of alloy specimens 2 to 9 from 90° to 293°K, is shown in figure 1. The mass susceptibility χ is throughout stated in e.m.u. g^{-1} for all figures.

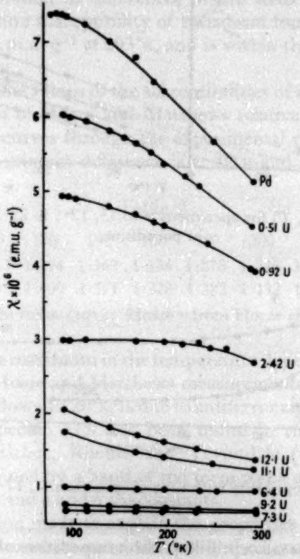


Figure 1. Temperature dependence of susceptibility of specimens 1 to 9, 10 to 12·08 atomic per cent uranium in palladium.

Specimens 8 and 10 to 14 obey a Curie–Weiss law of the form $\chi = C/(T-\theta)$ (with the exception of the low temperature behaviour of specimen 11), and in figure 2 the reciprocal of the mass susceptibility is plotted against the absolute temperature.

The susceptibility of specimen 11 was found to be time dependent below room temperature. On cooling the specimen to 90°K and holding it at that temperature, the susceptibility increased with time. On raising the temperature, after an anneal at 90°K, to temperatures above approximately 150°K, the susceptibility decreased with time, and the rate of decrease of susceptibility with time increased as the difference between 90°K and the higher temperature increased. Two quantitative measurements of this time variation were made. Before each of the measurements, the specimen was held at 90°K for three hours. Figure 3 shows the variation with time at 197°K and 262·5°K. Once equilibrium had been reached at these temperatures, no further variation with time was observed when the specimen was at any temperature between 197°K and 293°K.

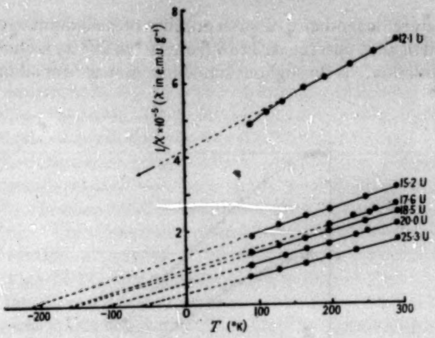


Figure 2. Curves of $(1/\chi, T)$ for specimens 9 to 15, 12·1 to 25·3 atomic per cent uranium in palladium.

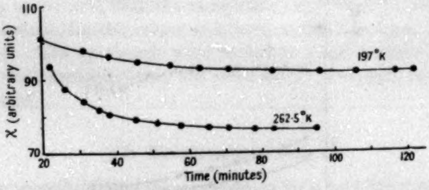


Figure 3. Change of susceptibility with time of specimen 11, 17·6 atomic per cent uranium in palladium.

The variation of mass susceptibility with temperature of the uranium specimen is shown in figure 4 (*b*). At 90°K the specimen showed a field dependence of susceptibility, and figure 4 (*a*) shows the graph of susceptibility plotted against the reciprocal of the field strength at 90°K, which was used to find the field independent susceptibility by the Honda and Owen method (cf. Bates 1951, p. 133). The field independent value at 90°K is plotted in figure 4 (*b*).

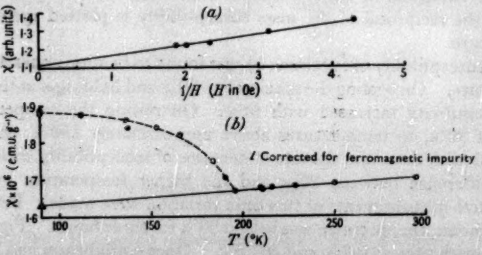


Figure 4 (*a*). Test for ferromagnetic impurity in uranium specimen at 90°K.
(*b*) Graph of (χ, T) of uranium specimen for field-independent values of χ.

§ 4. Discussion of Results

Accurate measurements on three specimens of spectroscopically standardized palladium were made by Hoare and Matthews (1953) who found values for the mass susceptibility of $5{\cdot}23\times10^{-6}$, $5{\cdot}32\times10^{-6}$ and $5{\cdot}13\times10^{-6}$ e.m.u. g^{-1} at 293°K and they attributed the variation to differences in the state of strain of the several specimens. The absolute susceptibility of palladium found in the present work was $5{\cdot}17\pm0{\cdot}04\times10^{-6}$ e.m.u. g^{-1} at 293°K, and is within the range of values found by Hoare and Matthews.

In table 2 are given the values of the susceptibilities of our palladium specimen and the third specimen of Hoare and Matthews relative to the value at 293°K, obtained from smooth curves through the experimental results at corresponding temperatures. The maximum difference is at 80°K and is approximately 1½%.

Table 2

T(°K)		80	100	120	140	160	180	200	220	240	260	280
$\frac{\chi_T}{\chi_{293}}$	mean	1·425	1·417	1·394	1·362	1·324	1·278	1·226	1·176	1·127	1·077	1·029
	Hoare	1·446	1·435	1·409	1·371	1·328	1·282	1·232	1·182	1·132	1·083	1·033

mean = from mean curve; Hoare = from Hoare *et al.* (1953).

It is unlikely that the maximum in the temperature dependence of susceptibility shown clearly in the Hoare and Matthews measurements of temperature dependence of susceptibility down to 20°K, is due to antiferromagnetic ordering, since no discontinuity in the specific heat has been found in this region (Clusius and Schachinger 1947). Elcock, Rhodes and Teviotdale (1954) suggest that the maximum can be explained by a band of the form $\nu(\epsilon)=a+b\epsilon^3$, where $\nu(\epsilon)$ is the energy density of states and a and b are constants.

The usual assumption made to explain the properties of a solid solution alloy is that the addition of the solute metal alters only the Fermi energy of the solvent, leaving the band form unaltered. The outer electronic configuration of palladium in the metallic state is $4d^{9\cdot4}\,5s^{0\cdot6}$ (Wohlfarth 1948), so that there is 0·6 hole per atom in the d band. The density of states at the Fermi level is high, partly explaining the large paramagnetic susceptibility of palladium. Hoare, Matthews and Walling (1953) examined the properties of a solid solution of silver in palladium, silver being a neighbour of palladium in the periodic table and having an electronic configuration $4d^{10}5s^1$. If the above assumption is correct, the ten outer electrons of palladium are replaced by the eleven outer electrons of silver and the extra electron enters the d band of palladium. At a concentration of 60 atomic per cent silver, Hoare, Matthews and Walling found that the alloy had a negligible susceptibility. This is the concentration corresponding to the filling of the d band on the above assumption. This assumption is discussed by Wohlfarth (1953) and applies to other palladium alloys.

Figure 5 shows the mass susceptibilities of our palladium–uranium alloys and of the palladium–silver alloys used by Hoare, Matthews and Walling (1953) plotted against the ‘ electronic concentration ’. The electronic concentration is defined as the atomic concentration for the silver alloys, and the atomic concentration multiplied by six for the uranium alloys. To correct approximately for the core diamagnetism a value of $0{\cdot}2\times10^{-6}$ e.m.u. g^{-1} has been added to the observed values of mass susceptibility.

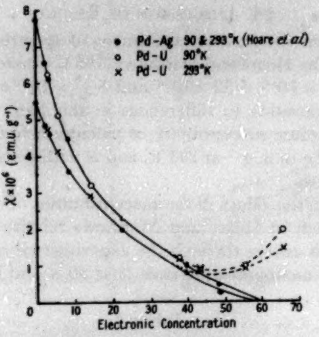


Figure 5. Mass susceptibilities of alloys plotted against electronic concentration.

The susceptibilities of both systems coincide up to a value of the electronic concentration of approximately 35. Making use of the previously stated assumption, it could be concluded that six outer electrons of uranium enter the d band of palladium. However, the assumption that the density of states is unaltered in a palladium–uranium alloy is probably not correct, since uranium has 46 more electrons than palladium and has an outer electronic configuration $5d^{10}\,5f^3\,6s^2\,6p^6\,6d\,7s^2$ (the number of 5f and 6d electrons is not known with certainty). Since the results appear to support an explanation based on the filling of the d band and we find that the assumption is doubtful in the case of uranium, although it is reasonable for neighbouring elements, one feels that the assumption may not provide the fundamental explanation for the behaviour of a metal dissolved in palladium. Probably all that can be stated from magnetic results alone is that the valency of the dissolved metal is an important factor.

At concentrations greater than 11 atomic per cent uranium, the susceptibility becomes strongly temperature dependent with a tendency to follow a Curie–Weiss law of the form $\chi = C/(T-\theta)$ (cf. figure 2) and begins to increase with increasing amounts of uranium seen from figure 6.

A possible explanation of this variation with temperature and concentration is that at higher concentrations the valency of uranium changes, leaving unpaired electrons attached to or localized on uranium atoms. The observed increase of susceptibility with concentration would then be due to these electrons. In order to test this explanation, it is necessary to know the probable valency of uranium and the spectroscopic state of the localized electrons. The metallurgical data on uranium and its alloys were reviewed by Mott (1952), who concluded that uranium has a valency of 4 or 6 in the metallic state. The valency of 6 has been used to explain the results with low concentrations of uranium, and if the valency changes to 4 at higher concentrations then the large observed paramagnetism could arise from the two remaining uranium electrons.

It is not possible to predict with certainty the spectroscopic state of the two electrons localized on uranium atoms, since the question of the precise position in the periodic table where the actinide starts has not been definitely answered (cf. Seaborg 1949, Haissinsky 1949). If a U^{4+} ion has two electrons in the 6d

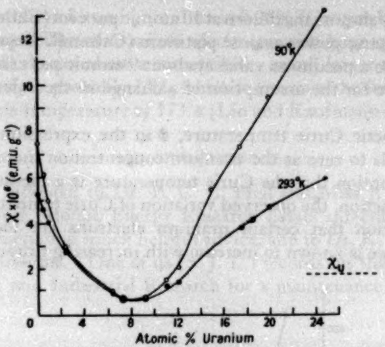


Figure 6. Susceptibility of palladium–uranium alloys plotted against concentration of uranium at 90°K and 293°K.

shell with orbital moment quenched, then the calculated effective atomic moment is 2·83 Bohr magnetons. If a U^{4+} ion has two electrons in the 5f band, then the calculated effective atomic moment is 3·58 Bohr magnetons.

In table 1 are given the values of the effective atomic moment for the uranium atoms calculated on the assumption that the magnetic contribution of the palladium atoms can be neglected. The value of the atomic moment for alloys of concentration greater than 15 atomic per cent is constant within experimental error, the mean value being 3·14 Bohr magnetons. This value is consistent with the assumption of a 4-valent state of the uranium atoms, but does not identify the shell which contains the two outer uranium electrons. The observed value for the effective atomic moment could arise from imperfect quenching of orbital motion in the 6d shell or from partial quenching in the 5f shell. The value of the atomic moment is in fair agreement with recently determined values of the effective atomic moments of uranium atoms in 4-valent salts which are listed in table 3.

Table 3. Effective Atomic Moment of U^{4+} in Uranium Salts

Salt	Atomic moment	Reference
Chloride	3·14	Stoenner & Elliott 1951
Chloride	3·29	Dawson 1951
Sulphate	3·52	Hutchison & Elliott 1948
Oxide	3·20	Dawson & Lister 1950
Oxalate	3·75	Hutchison & Elliott 1948
Fluoride	3·30	Elliott 1949
Fluoride	3·28	Dawson 1951
Bromide	3·12	Dawson 1951

At a concentration of approximately 10 atomic per cent uranium, several other properties of uranium–palladium alloys show a discontinuity. The variation of lattice parameter with concentration changes slope (Catterall, unpublished), the electrical resistivity shows a maximum at 10 atomic per cent (Catterall, unpublished) and the thermoelectric power against platinum (Catterall, unpublished) becomes positive and goes to a maximum value at about 9 atomic per cent. This provides additional evidence for the assumption of a change in the valency of uranium in this region.

The paramagnetic Curie temperature, θ in the expression $C/(T-\theta)$, of the alloys rapidly tends to zero as the uranium concentration increases (cf. figure 7). Making the assumption that the Curie temperature is a measure of the overlap of orbital wave function, the observed variation of Curie temperature is consistent with the assumption that certain uranium electrons are localized, since the interatomic distance is known to increase with increasing uranium content.

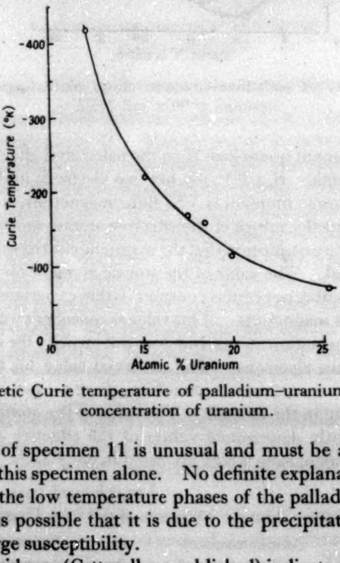


Figure 7. Paramagnetic Curie temperature of palladium–uranium alloys plotted against concentration of uranium.

The behaviour of specimen 11 is unusual and must be attributed to the heat treatment given to this specimen alone. No definite explanation for its behaviour can be given until the low temperature phases of the palladium–uranium system are known, but it is possible that it is due to the precipitation and solution of a second phase of large susceptibility.

Metallurgical evidence (Catterall, unpublished) indicates that quenched specimens containing up to 21 atomic per cent uranium could not contain significant quantities of UPd_3. Other intermetallic compounds UPd and U_5Pd_6 are formed only with considerable difficulty and do not come into equilibrium with the palladium-rich solid solution. This indicates that the large susceptibilities of solid solutions of uranium in palladium of concentration greater than 10 atomic per cent could not arise from the presence of increasing amounts of a second phase as an impurity but does not indicate any explanation of the behaviour of the annealed specimen.

The results of figure 4 (*b*) show that the temperature dependence of susceptibility of uranium below room temperature is a smooth extension of the high temperature results of Bates and Hughes (1954). The ferromagnetism arising at low temperatures is probably due to an impurity although it is unlikely that this is iron, since the solubility of iron in uranium is 0·06% (Grogan 1950) and measurements of the susceptibility of uranium specimens with large quantities of added iron (Mallard 1952) do not show a susceptibility much larger than that observed here. The impurity is probably UH_3 which is known to be ferromagnetic with a ferromagnetic Curie temperature of 173°K (Lin and Kaufmann 1955).

ACKNOWLEDGMENTS

We are much indebted to Dr. P. C. L. Pfeil and Miss J. R. Murray of the Metallurgy Division, Atomic Energy Research Establishment, Harwell, for the provision of specimens and much helpful advice, and to Dr. K. W. H. Stevens for many helpful discussions. One of us (S. J. L.) records his thanks to the Department of Scientific and Industrial Research for a maintenance grant held during part of this work.

REFERENCES

BATES, L. F., 1951, *Modern Magnetism*, 3rd edn (Cambridge : University Press).
BATES, L. F., BAKER, C. J. W., and MEAKIN, R., 1940, *Proc. Phys. Soc.* B, **52,** 182.
BATES, L. F., and HUGHES, D., 1954, *Proc. Phys. Soc.* B, **67,** 28.
BATES, L. F., and MALLARD, J. R., 1950, *Proc. Phys. Soc.* B, **63,** 520.
CLUSIUS, K., and SCHACHINGER, L., 1947, *Naturforsch.*, **2A,** 90.
DAWSON, J. K., 1951, *J. Chem. Soc.*, 429.
DAWSON, J. K., and LISTER, M. W., 1950, *J. Chem. Soc.*, 2181.
ELCOCK, E. W., RHODES, P., and TEVIOTDALE, A., 1954, *Proc. Roy. Soc.* A, **221,** 53.
ELLIOTT, N., 1949, *Phys. Rev.*, **74,** 498.
GROGAN, J. D., 1950, *J. Inst. Met.*, **77,** 571.
HAISSINSKY, M., 1949, *J. Chem. Soc.*, S 241.
HOARE, F. E., and MATTHEWS, J. C., 1953, *Proc. Roy. Soc.* A, **212,** 137.
HOARE, F. E., MATTHEWS, J. C., and WALLING, J. C., 1953, *Proc. Roy. Soc.* A, **216,** 502.
HUTCHISON, C. A., and ELLIOTT, N., 1948, *J. Chem. Phys.*, **16,** 920.
LEACH, S. J., 1955, *Thesis*, University of Nottingham.
LIN, S. J., and KAUFMANN, A. R., 1955, *A.I.E.E. Technical Conference on Magnetism*, June 1955.
MALLARD, J. R., 1952, *Thesis*, University of Nottingham.
MOTT, B. W., 1952, *A.E.R.E. Report* M/TN/18.
NETTLETON, H., and SUGDEN, S., 1939, *Proc. Roy Soc.* A, **173,** 313.
SEABORG, G. T., 1949, *Nucleonics*, **5,** 16.
STOENNER, R., and ELLIOTT, N., 1951, *J. Chem. Phys.*, **19,** 950.
WOHLFARTH, E. P., 1948, *Proc. Leeds Phil. Soc.*, **5,** 89; 1953, *Rev. Mod. Phys.*, **25,** 219.

The Magnetic Susceptibilities of Two Polymorphic Forms of Manganese Sulphide

By W. S. CARTER and K. W. H. STEVENS
University of Nottingham

MS. received 27th February 1956

Abstract. The magnetic susceptibilities of two polymorphic forms of MnS, one of zinc blende type and the other of wurtzite type have been measured. Antiferromagnetism occurs in the former at 100°K, with θ/T_c equal to 9·4. The results are discussed in the light of current views on antiferromagnetism and it is concluded that in the wurtzite type it may not be possible to form magnetic super-lattices. Some ordering will however occur, but it is not clear whether this can be detected magnetically. The considerations are extended and applied to the rare earth metals.

§ 1. Introduction

Manganese sulphide is known to crystallize in three polymorphic forms which are of considerable interest in studying the phenomenon of antiferromagnetism. The usual structure is of the sodium chloride type, in which each manganous ion is surrounded by a regular octahedron of sulphur ions and each sulphur ion is surrounded by a regular octahedron of manganous ions. The other two forms are more difficult to prepare and do not seem to have been investigated previously. Their structures are of the zinc blende and wurtzite types so that, in both, each manganous ion is surrounded by a regular tetrahedron of sulphur ions and each sulphur ion by a regular tetrahedron of manganous ions. The difference between the latter two structures lies in the arrangement of the tetrahedra; in the zinc blende type the manganous and sulphur lattices are both face-centred cubic, whereas in the wurtzite type the manganous and sulphur lattices are hexagonal with c/a almost the value for closest packing. The manganese–sulphur distances in the two structures are practically identical: 2·425 Å in the zinc blende type and 2·427 Å in the wurtzite type. In the sodium chloride type each manganous ion is known to have $S=5/2$ with $g\sim2$, and antiferromagnetism sets in at 165°K, with an asymptotic Curie temperature of 528°K. The latter values are given by Bizette (1946), and differ somewhat from those of Serres (1947). The type of ordering (the manganous lattice is face-centred cubic) is that described by Van Vleck (1951) as ordering of the second kind. Its essential feature is that the exchange interactions between next nearest neighbours are greater than those between nearest neighbours. The purpose of this paper is to give the results of our measurements on the other two forms and to discuss their implications.

§ 2. Experimental Details

All the results below were obtained from two samples prepared by Schnasse's (1932, 1933) method. It was not possible to obtain samples which were definitely of one type, and we had to be content with one sample which was mainly of wurtzite structure, but with a certain amount of zinc blende type, and another which was mainly of zinc blende type but with a certain amount of wurtzite type. Such mixing of the two structures is perhaps not surprising in view of the similarity between face-centred cubic and hexagonal close-packed structures and because the tetrahedra are practically identical in size in the two forms. Possibly both structures may occur in the same crystallite. The relative amounts of the two types were estimated by comparison of the intensities in the X-ray powder patterns with the calculated intensities for the two structures.

The first sample was estimated to contain 3·5 parts of the zinc blende type to one part of the wurtzite structure, and its variation of inverse susceptibility $1/\chi$ with temperature is shown by curve 1 of figure 1. The second specimen contained approximately four parts of the wurtzite structure to one part of the zinc blende type, and its $(1/\chi, T)$ curve is given by curve 2 of figure 1. For the measurement of relative susceptibility a Gouy method was used, with a maximum field strength of 4090 Oe. The force on the specimen was measured by means of an electrodynamic torsion balance based on a design due to Hutchison and Reekie (1946). The specimen, in the form of a fine powder, was contained in an evacuated Pyrex glass tube of 4 mm internal diameter, filling it to a length of 10 cm. The experimental arrangement was designed so that the glass tube passed through the magnetic field. These measurements are considered to be accurate to $\pm\frac{1}{4}\%$. The measurement of the absolute susceptibility and the ferromagnetic impurity test were kindly made at room temperature by Mr. R. G. Loasby. The absolute susceptibility is considered accurate to $\pm 2\%$ and with field strengths up to 9000 Oe there is no indication of ferromagnetic impurities.

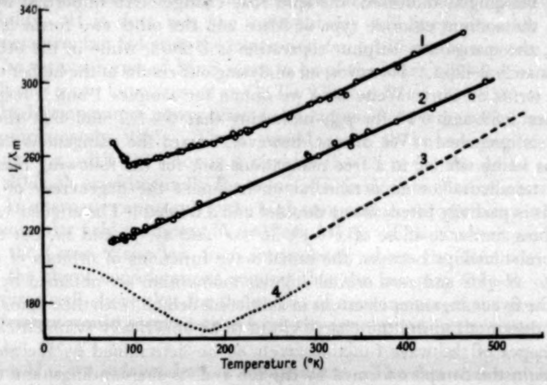


Figure 1. Variation in $1/\chi$ with T. 1, zinc blende type MnS; 2, wurtzite type MnS; 3, 4, NaCl type MnS.
(χ_m is the susceptibility in e.m.u. g mol of MnS.)

Measurements were made over the temperature range from the boiling point of liquid nitrogen to just below the temperature at which the wurtzite and zinc blende structures revert to the sodium chloride structure. The temperature was measured by means of a copper–constantan thermocouple attached to the specimen. This gave the temperature with an accuracy to ±2°K.

The parts of the curves below 110°K are the results of five temperature runs with the apparatus filled with hydrogen to prevent the condensation of oxygen on the specimen. The curves over the higher range are the results of three temperature runs. For clarity a number of experimental points in the vicinity of the minimum in (χ, T) for specimen 1 have been omitted from the curves. For the purpose of comparison the results of Serres (1947), curve 3, and Bizette (1946), curve 4, on the sodium chloride type of MnS are included in figure 1.

§ 3. Results and Discussion

In the two structures under consideration each manganese ion is surrounded by a tetrahedron of sulphur atoms, and one must first decide whether or not each ion has a spin of 5/2. This value is commonly obtained in the ionic manganous salts, but it is known that as the crystal field is increased or the tendency towards covalent bonding is increased, the spin may change. An important difference between the sodium chloride type of MnS and the other two forms is that, in the first, the manganese–sulphur separation is 2·606 Å, while in the others it is approximately 2·426 Å. However, on analysing our results at the higher temperatures in terms of Curie–Weiss laws we obtain for samples 1 and 2 respectively the values 5·89 and 6·19 for p_{eff}, indicating that $S=5/2$ and that the orbital moment is quenched. We do not, however, regard the manganous ion in the crystal as being similar to a free manganous ion, for the following reason. In either a tetrahedral or an octahedral environment the degeneracy of the five d-orbitals is partially lifted, into a doublet and a triplet. The angular variations of these are similar to those of x^2-y^2, $3z^2-r^2$ and xy, yz and zx, but there are now no relationships between the radial wave functions of orbitals of different energies. $S=5/2$ and zero orbital angular momentum are obtained by putting each of the five manganous electrons in a different orbital, with their spins parallel, but this does not require the charge cloud to be spherically symmetrical. The actual shapes of the wave functions seem to be determined by the amount of covalence in the complex formed by the ion and its surroundings, and this may be linked with the very different manganese–sulphur distances in the octahedrally and tetrahedrally co-ordinated groups, although some difference must be expected on purely electrostatic grounds. While perhaps not important in MnS, a further effect of the differences in radial wave functions is that the degeneracy of the ^{4}P level is lifted in a non-cubic environment (it is not lifted in first order using crystal field theory), and this can be used in understanding the fine structures observed in magnetic resonance on manganous ions. (A similar conclusion has been reached, using crystal field theory to higher orders in perturbation theory, by Dr. B. R. Judd (private communication).)

At approximately 100°K the susceptibility of sample 1 shows a discontinuity which we interpret as the onset of antiferromagnetism. Above this temperature the variation of $1/\chi$ with T is not quite linear, but using only the highest temperature points the susceptibility follows a Curie–Weiss law, $\chi=C/T+\theta$, with $\theta=938$°K. Sample 2 shows no sign of antiferromagnetism, and the $(1/\chi, T)$ curve is a straight line over the whole range of measurements, with θ again equal to 938°K. The values of these asymptotic Curie temperatures are much larger than the temperatures at which the measurements were made, and it is not obvious how they are to be interpreted. For, suppose that each manganous ion is interacting with every other ion and that the interaction energy can be written as ΣW_{ij}, where the suffixes i and j label the ions and W_{ij} is an interaction operator involving the spin components of i and j. Following the method of Van Vleck (1937) the susceptibility at high temperatures can be expressed as a power series in $1/T$:

$$\chi=\frac{A}{T}\left\{1+a_1\frac{1}{T}+a_2\left(\frac{1}{T}\right)^2+\ldots\right\} \qquad \ldots\ldots(1)$$

which can also be written as $\chi=A/(T+\theta)$, $(\theta \ll T)$, where θ depends on T but tends to the constant value $-a_1$ as T tends to infinity. Thus from high temperature measurements A and a_1 can be found, and the theory shows that A depends only on the moments of the ions, whereas a_1 depends on the W_{ij}. As the temperature is lowered the power series expansion is convergent until $1/T$ reaches a value equal to the radius of convergence of the series. What happens then depends upon whether or not the singularity lies on the positive real axis. If it does, then at the corresponding temperature there will be an anomaly in the susceptibility. If it does not, the susceptibility will vary smoothly through this temperature, but below it the expression for χ will not be a power series in $1/T$. The various forms of MnS seem to be of the latter type and we interpret the result that χ can be expressed as $C/(T+938)$ for T less than 938° as meaning that the high temperature power series has a radius of convergence corresponding to T approximately 938°. There would seem to be no justification for identifying $-a_1$ with 938°, nor is it obvious that A in (1) should be identified with C, although experimentally the value of C is close to that which we expect for A. An alternative approach to the high temperature susceptibility is provided by the molecular field treatment (Van Vleck 1951). This however predicts that χ obeys the Curie–Weiss law from immediately above the Néel temperature, whereas the above considerations suggest that the law may only be obeyed at temperatures well above that for the radius of convergence of the power series. This difference would seem to arise from the assumption in the molecular field theory that the α's and γ's etc. occurring in the molecular field expressions, $-\alpha M_A-\gamma M_B$, are independent of temperature. The fact that sample 1 does not give a straight line for the $(1/\chi, T)$ curve over the whole range above the Néel temperature can perhaps be explained in this way.

In sample 1 the lattice of manganous ions is face-centred cubic, and it might be expected that the ordering below the Néel temperature would conform with one of the types previously suggested for this lattice (Van Vleck 1951). In the absence of neutron diffraction results it is of interest to consider which type of ordering is most probable. Each sulphur ion is at the centre of a regular tetrahedron of manganous ions and it is therefore likely that the principal interactions are between nearest neighbour manganous ions, consisting of both direct exchange and exchange via the sulphur ions. The interactions between next nearest neighbours will probably be very small and, for the moment, can be neglected. If we suppose that the lowest energy state of two nearest neighbour manganous ions is when their spins are antiparallel, the lowest energy of a tetrahedral group is when, say, two are pointing east and two west, for this gives two energetically unfavourable and four energetically favourable pairs. With three spins pointing east and one west there are three unfavourable and three favourable pairs. We therefore assume that when ordering sets in it does so in such a way that each tetrahedron of manganous ions consists of two east and two west pointing spins. This is possible in many different ways. To obtain a super-lattice structure one must take account of next nearest neighbour interactions, weak though they may be. If they favour ferromagnetic alignment the super-lattice is that described as ordering of the first kind (Van Vleck 1951); if they favour antiparallel alignment it is improved ordering of the first kind. According to the molecular field theory the maximum possible value for θ/T_c is 3, compared with our experimental value of 9·4. We interpret this discrepancy as arising from inaccuracies in the molecular field theory, particularly when the next nearest neighbour interactions are weak.

In sample 2 the manganous lattice is hexagonal with c/a approximately equal to that for closest packing; otherwise the structure is very similar to that of sample 1, again consisting of tetrahedra of manganous ions about sulphur ions. We again expect that, in the ordered state, in each such tetrahedron two manganous ions will point east and two west. Ignoring next nearest neighbour interactions it is readily proved that there will be chains of ions, in the direction of the hexagonal axis, all with their spins parallel (a similar result can be proved for the body-centred cubic lattice, the chains being parallel to a cube edge). The chains must be arranged so that each manganous tetrahedron consists of two east and two west pointing spins, and again this is possible in many different ways. To determine whether a super-lattice can be formed it is necessary to include the next nearest neighbour interactions. It is fortunate that there are chains of similarly oriented spins parallel to the hexagonal axis, for this means that the problem can be examined in a projection upon a plane perpendicular to the hexagonal axis, thus reducing the three dimensional problem to one of two dimensions (see figure 2). Consideration of the patterns formed by an irregular arrangement of the chains, though such that all the tetrahedra consist of two east and two west pointing spins, shows that they can be regarded as a mosaic arrangement of rectangles containing four east (○) and two west (+) pointing spins, such a rectangle being shown in figure 3 (*a*). The question of the possibility of forming super-lattices under next nearest neighbour interactions then becomes equivalent to asking whether any particular arrangement of the rectangles is favoured. If this interaction favours parallel spins an obvious arrangement of the rectangles is that shown in figure 3 (*b*) for here all next nearest neighbours are parallel. When the next nearest neighbour interactions favour antiparallel spins it does not seem possible to have all next nearest neighbours antiparallel, and it is not at all clear that by making as many as possible favourable a super-lattice

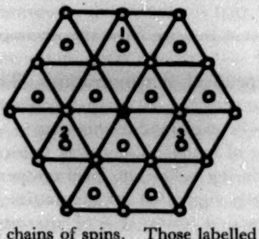


Figure 2. Projections of the chains of spins. Those labelled 1, 2 and 3 are next nearest neighbours of the one denoted by a filled-in circle.

is formed. If a super-lattice can be formed in this way we have not succeeded in finding it![†]

In view of the similarities in the tetrahedra in the zinc blende and wurtzite structures it seems likely that the ordering into chains will occur at some temperature. From our magnetic measurements it would seem to be below 100°K, although even this is not certain because we are not sure that the formation of the chains, without a super-lattice, can be detected magnetically. There is a clear need for further investigations of the wurtzite type of MnS, with which we are proceeding.

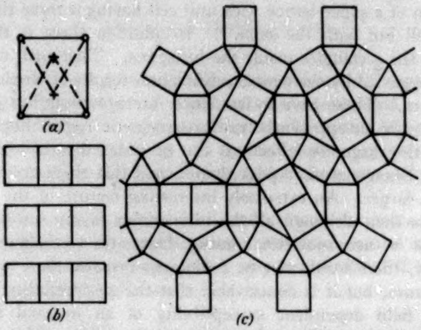


Figure 3. (*a*) shows a rectangular unit consisting of four east (○) pointing chains and two west (+) pointing chains. (*b*) shows an energetically favourable arrangement of rectangles when next nearest neighbours are ferromagnetically coupled. (*c*) shows a possible arrangement for antiferromagnetically coupled next nearest neighbours.

While the ideas discussed above were developed with MnS in mind, they have an obvious application in a relatively unrelated field. A number of the rare earth metals crystallize in structures which approximate to hexagonal close packed. Furthermore, they have localized magnetic moments. In MnS it was natural to discuss the antiferromagnetism in terms of the tetrahedra, and the same technique can be used with the rare earth metals. If we suppose that the interactions between neighbouring moments are such as to favour antiparallel alignments, the states of least energy occur when, in each tetrahedron, two spins point west and two east, for any change from this arrangement necessarily increases the energy. As we have seen above, this condition does not determine a super-lattice, nor do the next nearest neighbour interactions if they favour antiparallel alignments. However, the arguments used for MnS were based on the assumption that the next nearest neighbour interactions were much smaller than the nearest neighbour interactions. This may not be the case in a metal, where the exchange mechanism probably involves the conduction electrons. It is therefore conceivable, assuming both nearest and next nearest neighbour interactions favour antiparallel alignments, that a compromise is possible and some of the tetrahedra may not consist of two east and two west spins, if thereby more next nearest neighbours become antiparallel. An example of this is shown in figure 3 (*c*). The number of favourable next nearest neighbour interactions will be found on examination to be relatively high, but there are a few tetrahedra which contain one east and three west pointing spins. The symmetry of this figure would be hexagonal, were it not for the two rectangles, inside the hexagons, which can take any one of three positions. There is thus the strong suggestion of the formation of a super-lattice with unit cell having *a* three times that of the original unit cell but with the same *c*. In addition there is the randomness associated with the rectangles inside the hexagons. That such an arrangement may occur is suggested by the neutron diffraction results of Koehler and Wollan (1955) on erbium. These have so far defied interpretation, but it is interesting to see that there is an apparently antiferromagnetic region between 35°K and 80°K in which the magnetic reflections can be indexed on a unit cell which is three times the hexagonal close-packed unit, and that there are indications that it may be even larger. An extremely interesting feature of the 'super-lattice' of figure 3 (*c*) is that, although all the interactions favour antiferromagnetism, the lattice has a net moment coming from the unbalanced tetrahedra. Macroscopically, there would not be a moment because there would surely be a domain structure, but it is conceivable that the susceptibility would be field dependent. A field dependent susceptibility of an unusual kind has been observed in dysprosium (Elliott, Legvold and Spedding, 1954) between 176°K and 92°K. The super-lattice of figure 3 (*c*) is thus suggestive of a number of interesting phenomena which may be of importance in understanding the magnetic properties of the rare earth metals, though it must be remarked that these are so varied and unusual that further examination of them is required.

[†] Corliss, Hastings and Elliott (*Bull. Amer. Phys. Soc.*, 1956, **1**, 190, and BEL 2860) have reported the observation of super-lattices in the three forms of MnS. In the zinc-blende modification the ordering is ' improved ordering of the first kind ', in agreement with our prediction for antiferromagnetic next nearest neighbour interactions. The discovery of a super-lattice in the wurtzite structure has led us to re-examine our arguments, and we now find that a super-lattice can indeed be formed with antiferromagnetic next nearest neighbour interactions. It is identical with that observed.

ACKNOWLEDGMENTS

Our thanks are due to Professor L. F. Bates for the use of the facilities of his laboratory and for his help and encouragement at all times.

REFERENCES

BIZETTE, H., 1946, *Ann. Phys., Paris*, **1**, 295.

ELLIOTT, J. F., LEGVOLD, S., and SPEDDING, F. H., 1954, *Phys. Rev.*, **94**, 1143.

HUTCHISON, T. S., and REEKIE, J., 1946, *J. Sci. Instrum.*, **23**, 209.

KOEHLER, W. C., and WOLLAN, E. O., 1955, *Phys. Rev.*, **97**, 1177.

SCHNASSE, H., 1932, *Naturwissenschaften*, **20**, 641; 1933, *Z. Phys. Chem.* B, **20**, 89.

SERRES, A., 1947, *J. Phys. Radium*, **8**, 146.

VAN VLECK, J. H., 1937, *J. Chem. Phys.*, **5**, 320; 1951, *J. Phys. Radium*, **12**, 262.

Supersonic Dislocations and Dislocations in Dispersive Media

By J. D. ESHELBY
Department of Metallurgy, University of Birmingham

MS. received 23rd March 1956

Abstract. The conditions governing the motion of a screw dislocation at velocities above and below the velocity of shear waves are contrasted. For supersonic motion the Peierls–Nabarro equation becomes a differential equation, in contrast to the subsonic integral equation. It has a solution representing a moving restoration of fit along a slip plane across which there was originally complete misfit.

In a dispersive medium where the sound velocity decreases with decreasing wavelength a dislocation moving between the maximum and minimum sound velocities suffers a retarding force, which may be calculated if the shape of the dislocation is supposed to be known.

§ 1. INTRODUCTION

It is well known that the energy of a moving screw dislocation becomes infinite in 'relativistic' fashion as its velocity approaches the velocity c of transverse elastic waves (Frank 1949, Leibfried and Dietze 1949). There are, nevertheless, dislocation-like solutions of the elastic equations which move with velocities greater than c and which satisfy the Peierls condition or a prescribed generalization of it. A radiation of elastic energy away from the slip plane is always involved. For such supersonic velocities the Peierls–Nabarro equation becomes a differential equation in contrast to the integral equation of the subsonic case. The simplest solutions describe a crystal in which the atoms immediately above and below the slip plane start in a state of complete misfit; the passage of a supersonic ' dislocation ' then restores proper fit. These solutions may have some relation to the propagation of diffusionless transformations by dislocations.

In an actual crystal there is the complication of dispersion, the fact that the velocity of an elastic wave depends on its wavelength (and also on its direction and polarization). The sound velocity decreases from a maximum for infinite wavelengths to a minimum for the shortest wavelengths which the atomic structure allows. If it has a velocity below this minimum the dislocation, within the limitations of the elastic or Peierls–Nabarro model, moves freely. Above this minimum it is partly supersonic and so suffers a retarding force. The applied stress required to maintain its motion can be estimated if the width of the dislocation is known: disappointingly, it depends exponentially on the width. A complete investigation would involve setting up and solving the Peierls–Nabarro equation for the dispersive case. This is not attempted.

§ 2. SUBSONIC AND SUPERSONIC DISLOCATIONS

The elastic displacement w of a screw dislocation moving uniformly with velocity $v<c$ can be built up as a Fourier integral from elementary solutions of the wave equation

$$\frac{\partial^2 w}{\partial x^2}+\frac{\partial^2 w}{\partial y^2}-\frac{1}{c^2}\frac{\partial^2 w}{\partial t^2}=0 \qquad \ldots\ldots(1)$$

of the type

$$w=\exp\{i(x-vt)\mp\beta y\}, \quad \beta=(1-v^2/c^2)^{1/2} \qquad \ldots\ldots(2)$$

above and below the slip plane, that is, from plane waves moving parallel to the x axis and attenuated perpendicular to the direction of propagation. Stress and displacement on a plane $y=$ constant are related by

$$p_{zy}(x,y)=\mu\frac{\partial w}{\partial y}=\frac{\mu\beta}{\pi}\int_{-\infty}^{\infty}\frac{1}{\xi-x}\frac{dw(\xi,y)}{d\xi}\,d\xi.$$

In the model of Peierls and Nabarro the dislocation is held together by a ce law of force connecting stress and displacement on the atomic planes $y=$ adjacent to the slip plane, say

$$p_{zy}(x,\tfrac{1}{2}a)=-f[w(x,\tfrac{1}{2}a)]. \qquad \ldots\ldots(3)$$

The simple law

$$p_{zy}(x,\tfrac{1}{2}a)=-\frac{\mu b}{2\pi a}\sin\frac{4\pi w(x,\tfrac{1}{2}a)}{b} \qquad \ldots\ldots(4)$$

is satisfied by the displacement

$$w=\frac{b}{2\pi}\tan^{-1}\frac{\beta y}{x-vt}. \qquad \ldots\ldots(5)$$

The stress p_{zy} and velocity dw/dt [illegible] the elementary waves (2) are in quadrature, so that in any solution built up from them there is no net flux of energy perpendicular to the slip plane and the interatomic forces which give rise to the law (3) take care of the local energy transfers involved.

On the other hand if we look for a state of anti-plane strain moving uniformly with a velocity greater than c the wave equation takes the hyperbolic form

$$\frac{\partial^2 w}{\partial y^2}-\gamma^2\frac{\partial^2 w}{\partial(x-vt)^2}=0, \quad \gamma=i\beta=(v^2/c^2-1)^{1/2}$$

with the general solution

$$w=w_1(x-vt+\gamma y)+w_2(x-vt-\gamma y) \qquad \ldots\ldots(6)$$

i.e. plane waves of velocity c inclined to the x axis at such an angle that along a line $y=$ constant the profile of w propagates without change of form at a velocity v greater than c.

The solution (6) in the upper half-plane combined with the complementary solution

$$w=-w_1(x-vt-\gamma y)-w_2(x-vt+\gamma y)$$

in the lower half-plane will give a moving system of self-stress which could be maintained by a suitable law of force (3) across the slip plane. Physically, however, w_1 and w_2 are not on the same footing. They represent, respectively, plane waves moving obliquely out of and into the slip plane. Thus whereas w_1 can reasonably be said to be 'due to' the moving dislocation, the ingoing w_2 can only be maintained by elaborate arrangements at the outer boundary of the medium, and is better considered to be part of the externally applied elastic field. (In the electromagnetic analogy (Nabarro 1951, Eshelby 1953) [illegible] are associated with the retarded and advanced potentials of the ribbon of charge which replaces the displacement discontinuity across the slip plane.)

In particular for an infinite medium with a constant applied stress p^A,

$$w = \pm w_1(x - vt \pm \gamma y) + p^A y.$$

Since now the stress due to the dislocation is

$$p_{zy} = \mu \frac{\partial w_1}{\partial y} = \gamma\mu \frac{\partial w_1}{\partial x} = \gamma\mu \frac{\partial w}{\partial x}$$

the Peierls integral equation is replaced by the *differential* equation

$$\gamma\mu \frac{dw}{dx} + p^A = -f(w) \qquad \ldots\ldots(7)$$

in the supersonic case.

Suppose first that $p^A = 0$. Since p_{zy} and dw/dt are proportional there is everywhere a flux of energy out of the slip plane, and we might imagine that there would be no steady solution. But taking, for example, the law (4) there is the solution

$$w = \frac{b}{2\pi} \tan^{-1} \exp\{-2(x - vt)/\gamma a\}. \qquad \ldots\ldots(8)$$

w is zero before the passage of the disturbance, $\frac{1}{4}b$ afterwards. Comparison of (5) with figure 1(*a*) shows that $w = \frac{1}{4}b$ means perfect fit across the slip plane and $w = 0$ perfect misfit. (For clarity figures 1 (*a*) and (*b*) are drawn as if for an edge dislocation.) Let the upper half of the perfect crystal be shifted by $\frac{1}{2}b$ relative to the lower; then (8) evidently describes a restoration of fit progressing from left to right (figure 1 [illegible]).

Things are muc[illegible] same for an arbitrary law of force. We have

$$x - vt = -\gamma\mu \int_0^w \frac{dw}{f(w)}$$

and figure 1(*c*) shows that, as x goes from ∞ to $-\infty$, w swings from an unstable to a stable zero of f. The driving force to maintain the motion comes from the energy released as the region of misfit disappears. If there is a moderate applied stress we may introduce an effective force $f_1 = f(w) + p^A$. Then if p^A does not exceed the theoretical yield stress (the magnitude of the most negative value of f)

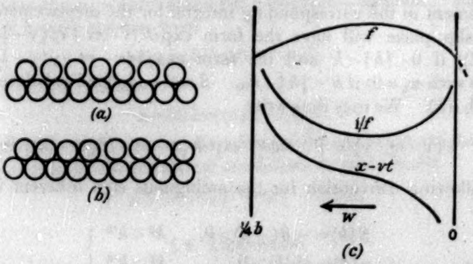


Figure 1

f_1 has zeros and figure 1(*c*) still applies, reading f_1 for f. For p^A greater than the theoretical yield stress f_1 is always positive and w is a linear function of $x-vt$ with a superimposed periodic ripple.

§ 3. THE EFFECT OF DISPERSION

Consider now a dispersive medium. The relation between frequency $\omega/2\pi$ and wave number k will have the general shape shown in figure 2. (We suppose that ω is independent of the direction of k.) $c(k)=\omega(k)/k$ will be a decreasing function of k. Let k_m be the largest value of k permitted by the atomic structure. Then a dislocation with velocity v will be subsonic with respect to all plane waves if a line of slope v on the diagram fails to cut the (ω, k) curve for k less than k_m. If, however, it intersects the curve (say at k'), it is partly subsonic, partly supersonic.

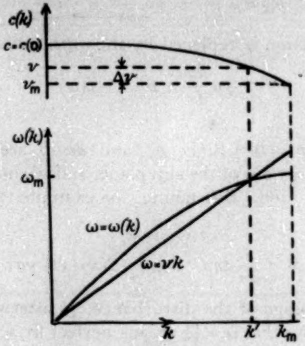


Figure 2.

If the displacement at the slip plane is

$$w(x-vt,\ \tfrac{1}{2}a)=\int_{-\infty}^{\infty} u(k)\exp ik(x-vt)\,dk$$

a typical element in the corresponding integral for the displacement $w(x-vt,y)$ above the slip plane will have the form $\exp ik\{x-vt+i\beta(y-\frac{1}{2}a)\}$ (compare equation (2)) if $0<|k|<k'$ and the form $\exp ik\{x-vt+\gamma(y-\frac{1}{2}a)\}$ (compare equation (6) with $w_2=0$) if $k'<|k|<k_m$. β and γ, defined as before, now depend on k through $c(k)$. We may thus write

$$w(x-vt,\ y)=\int_{-\infty}^{\infty} u(k)\ \exp ik\{x-vt+i\beta(y-\tfrac{1}{2}a)\}\,dk$$

with the following convention for the ambiguous sign inherent in the square roots:

$$\left.\begin{aligned}\beta(k)&=-\beta(-k)>0, && k^2<k'^2\\ \gamma(k)&=i\beta(k)>0, && k^2>k'^2.\end{aligned}\right\} \qquad \ldots\ldots(9)$$

The first condition ensures that the subsonic components fall off with increasing distance from the slip plane and the second that the supersonic components are outgoing waves.

The stress at the slip plane will be

$$p_{xy}(x, \tfrac{1}{2}a) = \int p(k) \exp(ikx)\, dk$$

with

$$p(k) = -\mu(k)\, k\, \beta(k)\, u(k)$$

where $\mu(k)$ is an effective elastic constant which we may take to be proportional to $c^2(k)$ for the supersonic components. Its value for the subsonic components will not be needed. (For brevity we have put $t=0$ and here and below we drop the limits of integration $\pm\infty$.) Similarly

$$\frac{dw}{dt}(x, \tfrac{1}{2}a) = \int q(k) \exp ikx\, dk$$

with $q(k) = -vik\, u(k)$. The flux of energy out of the slip plane is

$$E = -2 \int p_{xy} \frac{dw}{dt}\, dx$$

which by Parseval's theorem (Titchmarsh 1937) may be written

$$E = -4\pi \int p(k)\, q(-k)\, dk$$

or with the help of (9)

$$E = 8\pi v \int_{k'}^{k_{\mathrm{m}}} \mu(k)\, k^2 \gamma(k)\, u(k)\, u(-k)\, dk.$$

As we should expect, the subsonic components make no contribution. If v exceeds v_{m} only slightly, γ alone will vary appreciably over the range of integration; the other factors may be replaced by their value for $k = k_{\mathrm{m}}$. Moreover, since $\partial\omega/\partial k = 0$ at $k = k_{\mathrm{m}}$ we may put $c(k) = \omega(k)/k \simeq \omega_{\mathrm{m}}/k$ and

$$\gamma = [v/c(k) + 1]^{1/2}\, [v/c(k) - 1]^{1/2} \simeq 2^{1/2} (k - k')^{1/2} / k_{\mathrm{m}}^{1/2}.$$

Hence

$$E = (16\pi 2^{1/2}/3)\, v\mu(k_{\mathrm{m}})\, k_{\mathrm{m}}\, u(k_{\mathrm{m}})\, u(-k_{\mathrm{m}})\, (k_{\mathrm{m}} - k')^{3/2} / k_{\mathrm{m}}^{3/2}.$$

The final factor is $(\Delta v/v_{\mathrm{m}})^{3/2}$ where Δv is the excess of v over v_{m}. Then if, for definiteness, we give k_{m} and $\mu(k_{\mathrm{m}})$ the values π/b, $4\mu/\pi^2$ appropriate to the 100 direction in a simple cubic lattice of spacing b we have

$$E = (64\,.\,2^{1/2}/3b)\, v\mu\, u(k_{\mathrm{m}})\, u(-k_{\mathrm{m}}) (\Delta v/v_{\mathrm{m}})^{3/2}.$$

To find the form of $w(x)$ and hence $u(k)$ we should need to have a solution of the Peierls–Nabarro equation generalized for a dispersive medium. Failing this we suppose that w has the familiar form

$$w = (b/2\pi) \tan^{-1}(x/\zeta) \qquad \ldots\ldots(10)$$

with some value of ζ. Then

$$u = (b/2\pi) \exp(-\zeta\, |k|). \qquad \ldots\ldots(11)$$

E must be equated to bvp^{A}, the rate of working of the applied stress maintaining the motion. Rounding off the numerical coefficient we have finally

$$p^{\mathrm{A}}/\mu \simeq (\Delta v/v_{\mathrm{m}})^{3/2} \exp\{-2\pi\zeta/b\}. \qquad \ldots\ldots(12)$$

§ 4. DISCUSSION

The present calculations were originally undertaken in the hope that dispersion might provide a retarding mechanism which though unlikely to be the dominant effect, would at least provide an upper limit to the velocity of dislocations sufficiently unambiguous to answer some questions about their behaviour. (For a discussion of other dissipation mechanisms see Seeger (1955).) Unfortunately, the presence of the exponential factor in (12) prevents us from saying much more than that a non-zero retarding force sets in above v_{m}. This factor appears because the effect depends on the strength of those Fourier components (11) of the displacement (10) which have a wavelength of order b. Even if we could solve the appropriate dispersive Peierls–Nabarro equation precisely the result would be extremely sensitive to the law of force $f(w)$ assumed. There must evidently be a similar sensitive factor associated with any property of a dislocation in which the short wavelength components play the dominant role (e.g. the Peierls anchoring force). When the velocity of the dislocation is only slightly greater than v_{m} (so that $\Delta v/v_{\mathrm{m}} \ll 1$) the proportion of supersonic components in the Fourier spectrum of its elastic field is small. We might perhaps assume that the dislocation is, so to speak, mainly held together by the subsonic components. The width ζ is then fixed by the Peierls–Nabarro integral equation. With the law of force (4) this gives $p^{\mathrm{A}}/\mu \sim 10^{-3}$ for $\Delta v/v_{\mathrm{m}} \sim 1/10$. The exponential in (12) is the square root of that in Nabarro's (1951) expression for the Peierls force, and is equal to the exponential in the modification proposed by Huntington (1955).

The effect we have discussed is entirely analogous to the retardation of an electron by Čerenkov radiation when it moves through a dispersive dielectric, but with an important difference of detail. In the Čerenkov effect the electron may be treated as a point charge. Analogously we could formally consider a 'point dislocation', putting $\zeta=0$ in (12) to obtain the simple result $p^{\mathrm{A}}/\mu \sim (\Delta v/v_{\mathrm{m}})^{3/2}$. But in fact the atomic structure determines not only the dispersive properties of the medium, but also the width of the dislocation, and so introduces the exponential factor.

Returning to the case of a non-dispersive medium we have seen in § 2 that equation (7) purports to have as solutions all physically significant states of internal stress which can be supported by the law (4) and which move uniformly with velocity greater than c. However, Seeger (1953) has discussed certain 'oszillatorische Eigenbewegungen' which satisfy the wave equation (1) in the elastic region and the law (4) across the slip plane, but which are not solutions of (7). The displacement is just of the form (6) with $w_1(x) = w_2(x)$, so that there is an equal mixture of ingoing and outgoing waves. There is no net transfer of energy into or out of the slip plane. In general this state of affairs could only be maintained by an elaborate mechanism at the outer surface of the medium. In the special case of an infinite plate of suitable thickness the ingoing wave becomes the reflection of the outgoing. The surfaces would, however, have to be 'acoustically flat' even for the shortest wavelengths involved, that is, to within a fraction of an atomic spacing. Moreover, the possibility of maintaining the motion depends on the plate being infinitely long, for the disturbance can be loosely described as a procession of 'dislocations' each of which depends for the maintenance of its motion on the wake of all its predecessors multiply reflected by the surfaces of the plate. Of course solutions of the type (8) are open to the reverse objection: strictly they require a complete absence of ingoing waves. Nevertheless the motion might be expected to persist even in face of the largely unorganized ingoing waves resulting from internal scattering and reflection at the microscopically rough outer boundary, particularly as we now have in the misfit energy a definite driving force located at the slip plane.

Figure 1 (*b*) is reminiscent of a spreading stacking fault. If we take the movement of a particular atom from an unstable to stable position relative to its neighbours as representing a more complex energy-releasing rearrangement of atoms we may take our solution as symbolically representing the motion of a transformation dislocation moving in the boundary between two phases. (Compare for example, Bilby and Christian 1956.)

REFERENCES

BILBY, B. A., and CHRISTIAN, J. W., 1956, *Institute of Metals Monograph and Report Series*, No. 18, 121.

ESHELBY, J. D., 1953, *Phys. Rev.*, **90,** 248.

FRANK, F. C., 1949, *Proc. Phys. Soc.* A, **62,** 131.

HUNTINGTON, H. B., 1955, *Proc. Phys. Soc.* B, **68,** 1043.

LEIBFRIED, G., and DIETZE, H. D., 1949, *Z. Phys.*, **126,** 790.

NABARRO, F. R. N., 1951, *Proc. Roy. Soc.* A, **209,** 278.

SEEGER, A., 1953, *Z. Naturf.*, **8a,** 47; 1955, *Handbuch der Physik*, VII/1, p. 622 (Berlin : Springer).

TITCHMARSH, B. C., 1937, *Fourier Integrals* (Oxford : Clarendon Press).

The Statistics of Charge Carrier Fluctuations in Semiconductors

By R. E. BURGESS†

Radio Research Station, Slough, Bucks.

Communicated by R. L. Smith-Rose; MS. received 28th November 1955,
and in revised form 26th March 1956

Abstract. The statistical fluctuations in the number of charge carriers (conduction electrons or valence band holes) in a given volume of semiconductor in the steady state is calculated from the transition probabilities of electrons between the conduction and valence bands or between either band and impurity levels. The case in which one independent fluctuating variable (either n or p) specifies the electronic state of the system is considered. Simple general formulae are developed for the mean value n_0 of n and its variance and it is shown that for large numbers n tends to be normally distributed about n_0. The relaxation time for small deviations from equilibrium is evaluated. These results are applied to three cases: (i) the intrinsic semiconductor, (ii) the strongly extrinsic semiconductor and (iii) the slightly extrinsic semiconductor with all impurity atoms ionized. The relation between the statistical approach and the thermodynamical treatment (given in another paper) is indicated.

§ 1. INTRODUCTION

THE number of conduction electrons or of holes in the valence band in a given volume V of semiconductor undergoes spontaneous fluctuations due to the thermal excitation of carriers and their recapture; these fluctuations are of interest in the theory of noise generation in semiconductors. The average number of carriers and the mean square fluctuation of the number about the average can be calculated if the various transition probabilities of electrons between the bound levels, the conduction band and the valence bands are known. These probabilities depend upon the existing degrees of electron occupancy of the bands and the bound levels and fluctuate as the occupancy fluctuates. It is assumed only that the system is in a steady state but not necessarily in thermal equilibrium.

In the present analysis it will be assumed that there is only one independently fluctuating variable which may be either the number of holes p in the valence band or the number of electrons n in the conduction band. This single variable then determines also the number of electrons in the other relevant band or levels. This simple case enables three important cases of a neutral non-degenerate semiconductor to be considered: (i) the intrinsic semiconductor (in which $n=p$) in which the transitions are between the valence and conduction bands and no bound levels are involved, (ii) the strongly extrinsic semiconductor in which the number of minority carriers can be ignored and the transitions take place between the impurity levels and either the conduction or valence bands, (iii) the slightly extrinsic semiconductor in which all the N impurity atoms are completely ionized and the transitions occur between the conduction bands and either the valence band or electron traps.

---

† Now at Department of Physics, University of British Columbia, Canada.

§ 2. General Analysis

When there are n electrons in the conduction band let the probability that another electron enters the band in time dt be $g(n)dt$ and let the probability that an electron leaves the band during dt be $r(n)dt$. It is assumed that the generation and recombination rates g and r depend on only one fluctuating variable, in this case n. In the steady state the probability distribution function P of n will be stationary and this leads to the functional form of $P(n)$:

$$\frac{P(n)}{P(0)}=\prod_{\nu=0}^{n-1} g(\nu)\Big/\prod_{\nu=1}^{n} r(\nu)$$

as a consequence of the steady state condition

$$0=\frac{dP(n)}{dt}=r(n+1)P(n+1)+g(n-1)P(n-1)-P(n)[g(n)+r(n)].$$

The most probable value n_0 of n may be found by logarithmic differentation of P and equating the derivative to zero:

$$0=\frac{d}{dn}\ln P(n)=\ln\frac{g(n_0)}{r(n_0)}.$$

It is assumed throughout that all the numbers n, p and N are large compared with unity. The equilibrium condition corresponds to equal rates of recombination and generation as expected:

$$g(n_0)=r(n_0). \qquad \ldots\ldots(1)$$

By expressing ln $P(n)$ as a quadratic expansion about its maximum value ln $P(n_0)$ we are led to the corresponding gaussian distribution of n about n_0:

$$\left[\frac{d^2}{dn^2}\ln P(n)\right]_{n=n_0}=\frac{g'(n_0)}{g(n_0)}-\frac{r'(n_0)}{r(n_0)}$$

where the prime denotes differentation with respect to n.

Thus $$\ln P(n)=\ln P(n_0)-\tfrac{1}{2}(n-n_0)^2\left[\frac{r'(n_0)}{r(n_0)}-\frac{g'(n_0)}{g(n_0)}\right]$$

or $$P(n)=P(n_0)\exp\left[-(n-n_0)^2/\overline{2(n-n_0)^2}\right] \qquad \ldots\ldots(2)$$

where the variance of n is given by the very simple formula

$$\overline{(n-n_0)^2}=\left[\frac{r'(n_0)}{r(n_0)}-\frac{g'(n_0)}{g(n_0)}\right]^{-1}=\frac{g(n_0)}{r'(n_0)-g'(n_0)}. \qquad \ldots\ldots(3)$$

The problem has been formulated in a general manner and is therefore applicable to a variety of physical systems, in which g and r can be regarded as transition probabilities in a reversible process such as the thermal excitation of carriers, the diffusion of electrons, atoms or lattice defects in a crystal, or the statistics of the interchange of energy between photons and a material system.

A few general remarks may be offered about the fluctuation formula. Usually the rate of generation g will be either constant or a decreasing function of n while r will be an increasing function of n; these conditions incidentally ensure that the denominator in equation (3) shall be positive. In many cases of practical interest one may regard the generation function as a linearly decreasing function of n while the recombination function is a quadratic increasing function of n:

$$g(n)=g_0-an \qquad r(n)=bn+cn^2 \qquad \ldots\ldots(4)$$

where the coefficients g_0, a, b and c are independent of n. Several specific cases which conform to this pattern are discussed later.

The equilibrium value of n will be given by

$$n_0 = -\frac{a+b}{2c} + \left[\left(\frac{a+b}{2c}\right)^2 + \frac{g_0}{c}\right]^{1/2} \qquad \ldots\ldots(5)$$

and the variance of n will be

$$\overline{(n-n_0)^2} = n_0 \frac{b+cn_0}{a+b+2cn_0} \qquad \ldots\ldots(6)$$

which can never exceed n_0 and must lie between $n_0 b/(a+b)$ and $\frac{1}{2}n_0$.

Another aspect of the fluctuation problem is the relaxation time determining the return to equilibrium after a small deviation from equilibrium. If at any given moment $n \neq n_0$ the expectation of the rate of change of n towards n_0 is

$$\overline{\frac{d}{dt}(n-n_0)} = \frac{\overline{dn}}{dt} = g(n) - r(n) \simeq -[r'(n_0) - g'(n_0)](n-n_0)$$

so that for small deviations the relaxation time is

$$\tau = 1/[r'(n_0) - g'(n_0)]. \qquad \ldots\ldots(7)$$

§ 3. Special Cases

3.1. *Intrinsic Semiconductor*

In a non-degenerate intrinsic semiconductor (figure 1(a)) it may be assumed that the rate of generation of hole–electron pairs is proportional to the number of un-ionized semiconductor atoms $N-n$ which may be taken as constant since N

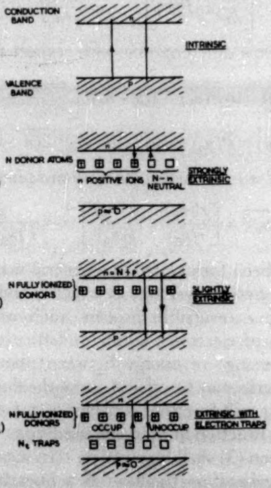

CONDUCTION BAND
(a)
INTRINSIC
VALENCE BAND
n
p
N DONOR ATOMS
n POSITIVE IONS
N−n NEUTRAL
(b)
STRONGLY EXTRINSIC
p≈0
n=N+p
N FULLY IONIZED DONORS
(c)
SLIGHTLY EXTRINSIC
N FULLY IONIZED DONORS
OCCUP
UNOCCUP
(d)
Nt TRAPS
EXTRINSIC WITH ELECTRON TRAPS
p≈0

Figure 1. Energy band system and electron transitions involved in the fluctuations of conduction electrons in an n-type semiconductor.

will be very much larger than n; the rate of recombination is proportional to the product of the number of electrons and holes, i.e. to n^2 since $n=p$ in an intrinsic semiconductor. Thus we have

$$g=\text{constant}, \qquad r=\rho n^2 \qquad \ldots\ldots(8)$$

where ρ is a parameter independent of n. The equilibrium value of n is thus

$$n_0=(g/\rho)^{1/2}\equiv n_1. \qquad \ldots\ldots(9)$$

In this case the variance of n is simply

$$\overline{(n-n_0)^2}=\tfrac{1}{2}n_1=\overline{(p-p_0)^2} \qquad \ldots\ldots(10)$$

and the relaxation time is $\tau=1/2\rho n_1$.

3.2. *Strongly Extrinsic Semiconductor*

If in an n-type semiconductor there are N donor atoms and n electrons in the conduction band and the number of holes p is sufficiently small compared with n to be negligible (figure 1(*b*)) the rate of electron excitation will be proportional to the number of neutral donor atoms $N-n$ while the rate of recombination will be proportional to the product of the number of conduction electrons and the number of positive donor ions, i.e. to n^2. Hence

$$g=\gamma(N-n), \qquad r=\rho n^2 \qquad \ldots\ldots(11)$$

but by contrast with the intrinsic case N is not in general large compared with n.

Then the equilibrium number of conduction electrons (given by $g=r$) is

$$n_0=-\tfrac{1}{2}K+(\tfrac{1}{4}K^2+KN)^{1/2} \qquad \ldots\ldots(12)$$

where $K=\gamma/\rho$ is the equilibrium constant for the ionization process. The mean square fluctuation of n about n_0 is given by

$$\overline{(n-n_0)^2}=\frac{n_0^2}{2n_0+K}=\frac{n_0(N-n_0)}{2N-n_0}=\left[\frac{2}{n_0}+\frac{1}{N-n_0}\right]^{-1} \qquad \ldots\ldots(13)$$

which is always less than $\frac{1}{2}n_0$. The relaxation time is given by $\tau=1/(\gamma+2\rho n_0)$.

Thus if the donors are almost completely ionized ($n_0\sim N$) as is the case when $K\gg N$ the variance tends to $n_0^2/K=N-n_0$; this implies that the variance of the number of un-ionized donor atoms $N-n$ is equal to the mean number. If the donors are only partially ionized as is the case when $N\gg K$ then $n_0\simeq(KN)^{1/2}$ and the variance of n tends to $\frac{1}{2}n_0$ just as in the case of the intrinsic semiconductor.

With acceptors instead of donors and p replacing n, the same arguments apply to a strongly extrinsic p-type semiconductor.

3.3. *Extrinsic Semiconductor with Fully Ionized Impurity Atoms*

If the Fermi level is sufficiently low in an n-type semiconductor that all the N donors can be regarded as ionized (figure 1(*c*)) and so long as the number of neutral donors can be neglected by comparison with the number of holes, the fluctuations due to the donor ionization may be ignored. In this case the number of electrons in the conduction band is $n=N+p$. We then consider only transitions between the valence and conduction bands as producing fluctuations whence

$$g=\text{constant}, \qquad r=\rho np=\rho n(n-N) \qquad \ldots\ldots(14)$$

where $g/\rho = n_i^2$ with n_i equal to the number of electrons in an intrinsic sample.

The equilibrium number of electrons is thus

$$n_0 = \tfrac{1}{2}N + (\tfrac{1}{4}N^2 + n_i^2)^{1/2} = N + p_0 \qquad \ldots\ldots(15)$$

and $n_0 p_0 = g/\rho = n_i^2$ as expected.

The variance in n or p is

$$\overline{(n-n_0)^2} = \overline{(p-p_0)^2} = \frac{n_0(n_0-N)}{2n_0-N} = \frac{n_0 p_0}{n_0+p_0} \qquad \ldots\ldots(16)$$

which reaches its maximum value of $\frac{1}{2}n_i$ in the intrinsic case ($N=0$). If n_0/p_0 greatly exceeds unity the variance approaches p_0. The relaxation time also has a symmetrical form in n_0 and p_0, viz. $\tau = 1/\rho(n_0+p_0)$.

This case is of particular application to nearly intrinsic material in which the Fermi level is nearly midway between the conduction and valance bands so that, in material in which the energy gap is not too small, all the donor atoms located just below the bottom of the conduction band can be regarded as fully ionized.

Similar results hold for a slightly p-type semiconductor in which N is the number of acceptor atoms (fully ionized) and n and p are interchanged. The symmetrical form for the variance of the number of carriers (n or p) $\overline{(n-n_0)^2} = \overline{(p-p_0)^2} = n_0 p_0/(n_0+p_0)$ still retains its validity. It may then be asserted that the variance is equal to the number of minority carriers, when n_0/p_0 departs appreciably from unity so long as fluctuations in the ionizations of impurity atoms can be ignored.

3.4. *Extrinsic Semiconductor with Electron Traps*

This case is similar to the last one but in practice it will be assumed that the transitions causing fluctuations in n are between the conduction band and electron traps and not between the conduction and valence bands (figure 1(*d*)). It will be further assumed that the number of holes is negligible. If N_t is the number of traps and N is the number of fully ionized donors there are two cases to distinguish:

(i) the traps are neutral when occupied and hence the number of occupied traps is $N_{t1} = N_t + N - n$.

Then

$$g = \gamma(N_t + N - n), \qquad r = \rho n(n-N) \qquad \ldots\ldots(17a)$$

$$\text{giving} \quad \overline{(n-n_0)^2} = \left[\frac{1}{n_0} + \frac{1}{N_{t1}} + \frac{1}{N_t - N_{t1}}\right]^{-1} \qquad \ldots\ldots(18a)$$

(ii) the traps are singly charged negative ions when occupied and hence the number of occupied traps is $N_{t1} = N - n$. Then

$$g = \gamma(N-n), \qquad r = \rho n(N_t - N + n) \qquad \ldots\ldots(17b)$$

$$\text{giving} \quad \overline{(n-n_0)^2} = \left[\frac{1}{n_0} + \frac{1}{N_{t1}} + \frac{1}{N_t - N_{t1}}\right]^{-1}. \qquad \ldots\ldots(18b)$$

The variances of n may be thus expressed in the same way for the two cases; further the variance in n is also the variance of N_{t1} the number of occupied traps. If the number of traps is small compared with the number of donor atoms the variance becomes of the form appropriate to the binomial law

$$\overline{(N_{t1} - N_{t10})^2} = \overline{(n-n_0)^2} = N_t f(1-f) \qquad \ldots\ldots(19)$$

which has a maximum value of $\frac{1}{4}N_t$ when the fractional occupancy $f(=N_{t1}/N_t)$ of the traps equals $\frac{1}{2}$.

The relaxation time when $N_t \ll N$ is given by

$$\tau = \frac{1}{\gamma + \rho N} = \frac{1-f}{\gamma} = \frac{f}{\rho N} \qquad \ldots\ldots(20)$$

and this will always be less than the full recombination time $1/\rho N$ by the trap occupancy factor f.

§ 4. Discussion

The variances derived in the preceding section (equations (10), (13) and (16)) are all in agreement with those derived by thermodynamical arguments using the second derivatives of the free energy (Burgess 1955). Similarity may be noted between the purely statistical treatment given here and the treatment involving the electronic free energy. When an electron makes a transition in a semiconductor (at constant volume and temperature) between one level and another with an energy difference ϵ (so increasing n by 1) there is a change of free energy ΔF of the system given by $\epsilon - T\Delta S$ where ΔS is the change of entropy due to the transition which in general will depend upon the number of electrons which are occupying the two levels at the time of the transition. If ϵ is regarded as independent of the numbers of electrons it is seen that

$$\frac{\partial^2 F}{\partial n^2} = -T\frac{\partial^2 S}{\partial n^2}.$$

If now we take $\partial^2 S/\partial n^2$ as given by $(\partial^2/\partial n^2)[\boldsymbol{k} \ln P(n)]$ where $\boldsymbol{k}$ is Boltzmann's constant, which implies that S and $\boldsymbol{k} \ln P(n)$ differ only by an additive constant and terms linear in n, we have

$$\frac{\partial^2 F}{\partial n^2} = -\boldsymbol{k}T\frac{\partial^2}{\partial n^2}\ln P(n).$$

Thus if F_0 is the equilibrium or minimum value of the electronic free energy occurring when $n = n_0$

$$\begin{aligned} F - F_0 &= \tfrac{1}{2}\left(\frac{\partial^2 F}{\partial n^2}\right)_{n_0}(n-n_0)^2 \\ &= \tfrac{1}{2}\boldsymbol{k}T(n-n_0)^2\left[-\frac{\partial^2}{\partial n^2}\ln P(n)\right]_{n_0}. \end{aligned}$$

The exponential distribution of $F - F_0$ namely $\exp[-(F-F_0)/\boldsymbol{k}T]$ leads to the gaussian distribution of n about n_0 with a variance of $[-(\partial^2/\partial n^2)\ln P(n)]_{n_0}{}^{-1}$ as derived in §2. Thus the formal correspondence between the two approaches is established. The effect of spin in the impurity levels (in the case of the strongly extrinsic semiconductor) is to increase the number of ways in which the electrons may be arranged in the impurity levels; this depends upon whether they have paired or unpaired electrons and introduces a term of $\boldsymbol{k}n \ln 2$ or $\boldsymbol{k}(N-n) \ln 2$ respectively into the entropy; this influences the equilibrium value n_0 but not the variance in n since the second derivative of this term vanishes. This effect could be included in the statistical method of this paper by distinguishing between + and − spins and letting $P(n, \nu)$ refer to the probability of n electrons in the conduction band with certain arrangements of electron spin in the donor levels defined by the number ν.

In the literature on noise in semiconductors there sometimes occur inappropriate expressions for the fluctuations in the number of electrons (or holes) due to faulty application of statistical theory. Thus in the case of electrons coming from bound levels (e.g. donors or traps) it is often assumed that the binomial distribution applies to the number of conduction electrons, i.e. if N is the total number of bound levels providing electrons and n is the number of conduction electrons present at any moment, it is asserted that

$$P(n)=\binom{N}{n}p^n(1-p)^{N-n}$$

where p is the probability that a bound level shall be unoccupied by an electron. This leads to the mean value and variance of n given by

$$n_0=pN \quad \text{and} \quad \overline{(n-n_0)^2}=Np(1-p).$$

Now the binomial $P(n)$ may be examined in terms of the generation and recombination rate functions which lead to it by noting that

$$\frac{g(n)}{r(n+1)}=\frac{P(n+1)}{P(n)}=\frac{p}{1-p}\,\frac{N-n}{n+1}$$

so that we can write $g(n)=\gamma(N-n)$, $r(n)=\gamma n(1-p)/p$ which corresponds to a generation rate proportional to the number of occupied levels $N-n$, and depends on a recombination rate directly proportional to the number of conduction electrons. This linear type of recombination does not arise under conditions of simple impurity levels for it is the bimolecular or quadratic type of recombination which will there apply (§ 3.2); however it may occur when there are trapping levels in addition to saturated impurity levels (§ 3.4).

In different circumstances it has sometimes been assumed that Poisson's law applies to n, namely in intrinsic or nearly intrinsic semiconductors, for which the electron fluctuations are assumed to be entirely or almost entirely due to fluctuations in the transitions between the valence and conduction bands. Now the Poisson distribution

$$P(n)=\frac{n_0{}^n\exp(-n_0)}{n!} \quad \text{gives} \quad \frac{g(n)}{r(n+1)}=\frac{P(n+1)}{P(n)}=\frac{n_0}{n+1}$$

corresponds to a constant rate of generation and a linear recombination rate proportional to n. Again it is the hypothesis of linear rather than quadratic recombination which leads to an erroneous value for the variance.

The intrinsic semiconductor although apparently the simplest system to consider may in fact be more complicated than the discussion of § 3 would suggest. This is because, at least in the case of germanium, the process of generation and recombination of hole–electron pairs is far more likely to occur via the trapping levels associated with crystal imperfections (especially at the crystal surface) than by transitions directly between the conduction and valence bands. This leads to some complexity in the statistics of the fluctuations since we can no longer define the electronic state of the system in terms of one variable (e.g. n) since n, p and the number of occupied trap levels are all fluctuating and are partially correlated; thus two fluctuating parameters must be handled and these can conveniently be taken as n and p while the condition of overall neutrality then determines the trap occupancy. When this type of system is considered it seems somewhat simpler to use the thermodynamical approach and in another paper (Burgess 1955) the author has treated the problem of simultaneous fluctuations in n, p and bound levels by this method.

It may be noted that the general methods developed by Fowler (1936) for calculating the equilibrium state of an assembly and the fluctuations about that state can be applied to the case of electron–ion systems here under consideration. Especially of interest in this connection is § 20.31 of Fowler's book leading to his equation (2146) for the variance of the number of particles in a classical dissociating assembly governed by a single reaction.

ACKNOWLEDGMENT

The work described above was carried out as part of the programme of the Radio Research Board. This paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

REFERENCES

BURGESS, R. E., 1955, *Proc. Phys. Soc.* B, **68**, 661.
FOWLER, R. H., 1936, *Statistical Mechanics*, 2nd edn (Cambridge : University Press).

The Freezing of Water Droplets

By A. E. CARTE

National Physical Laboratory, South African Council for Scientific and Industrial Research,
Pretoria, South Africa

MS. received 27th February 1956, *and in revised form 23rd April* 1956

Abstract. A microscope method was used to study the freezing of individual water droplets 9 to 33 microns in diameter formed by condensation on various surfaces and protected with a film of silicone oil. The lowest freezing temperatures were found to be independent of the six surfaces used but dependent upon the volume of the water droplets and the rate of cooling. The results were correlated with those of other observers and it was concluded that homogeneous nucleation occurred in the lowest freezing droplets. Values of the surface free energy for a water-ice interface were calculated.

§ 1. INTRODUCTION

STUDIES of the freezing of supercooled water date back to those of Fahrenheit in 1732. Since then many data on this subject have accumulated, especially during the last decade when the freezing of small droplets has received increasing attention. The need for information on the factors governing the freezing of cloud droplets has been stimulated, as have other branches of cloud physics, by experiments to control precipitation artificially. However, even now there are gaps in the knowledge. For example, almost a century ago Mousson (1858) found that smaller drops of water could be supercooled more than larger ones and yet the exact relationship between freezing temperature and volume is not known with certainty. Also it is not known whether water droplets enclosed by solid or liquid interfaces freeze at the same temperature as those in air; nor has the effect of cooling rate on freezing been unequivocally established.

The latest work in this field has been summarized by Mossop (1955) and Jacobi (1955).

§ 2. EXPERIMENTAL METHODS

Water droplets were formed in the presence of still, atmospheric air on the lower, highly polished surface of a cylindrical metal block one inch in diameter and 1¼ inches long inserted into a Perspex holder which fitted rigidly on to the stage of a metallurgical microscope (figure 1). A cylindrical Perspex cup, having a side tube, fitted on to the objective lens. A few drops of distilled water in this cup supplied water vapour for the dew which formed on the polished surface when the metal block was cooled. As soon as the droplets had grown to the required size, those located on the centre of the surface were covered with a drop of silicone oil (Midland Silicones 200 fluid, viscosity 1000 cs) and a glass cover slip, the latter spaced from the surface by means of a thin mica ring. Thus protected by the film of silicone oil, even the smallest droplets persisted for days thereby allowing the same droplets to be frozen and melted many times.

Temperatures were measured with 42 s.w.g. nichrome–eureka butt-welded thermocouples that had been calibrated against standard instruments. The particular thermocouple in use was inserted into a fine hole drilled along a diameter close to the polished face of the metal block, with the junction at the centre of the face and the wires insulated from the metal by means of a glass sleeve. To obtain the temperature of the droplets on the surface with this arrangement it was necessary to apply small corrections to compensate for heat conduction along the thermocouple leads and for lag effects at high rates of change of temperature. These corrections were found from observations on the melting at different rates of heating of mercury droplets and of water droplets on the surface. The accuracy of the temperature measurements is estimated to be to ±0·2°c, while differences under the same experimental conditions are significant to within ±0·1°c.

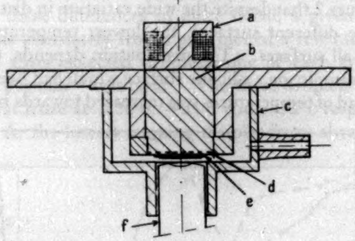


Figure 1. Apparatus for viewing the freezing of water droplets. a, heater; b, metal block; c, Perspex cylinder; d, thermocouple hole; e, polished surface; f, microscope objective lens.

Cooling of the metal block was effected through contact of its upper surface with the end of a copper hoop, the remote end of which dipped in liquid oxygen. It was more convenient for the higher cooling rates to seat a metal cylinder having insulated sides and filled with dry ice and acetone on top of the block. Cooling and heating were controlled by an electrical heater wound on to a brass bobbin between the block and the cylinder.

The condensation of moisture on the outside surface of the cover slide was prevented by maintaining an atmosphere of dry oxygen in this region. It was found best to maintain a very slow stream of oxygen, as water vapour from the droplets on the metal surface diffused through the silicone oil to the atmosphere, the process being speeded up as the oxygen stream was increased. This diffusion of water vapour through the silicone oil takes place from liquid to frozen water proplets. The process is slow but nevertheless does limit the number of freezes that can be carried out on a given field of droplets because those that freeze first grow at the expense of the others.

Simultaneous observations on a number of droplets were made photographically.

§ 3. THE EFFECT ON FREEZING POINT OF THE NATURE OF SURFACE UPON WHICH DROPLETS WERE FORMED

Water droplets were formed on six different surfaces, four metal and two non-metal, to find out how freezing temperatures are affected by the nature of the solid surface. Gold and silver surfaces were formed by evaporation *in vacuo* of these metals on to brass blocks, nickel by electro-plating and subsequent polishing and stainless steel by electrolytic polishing. The non-metals used were collodion and silicone. The silicone was baked on to a thin layer of silicon monoxide which had been deposited on to polished stainless steel.

The freezing of some hundreds of droplets 10 to 20 μ in diameter was observed on every surface. Water showed some tendency to wet the collodion and stainless steel. On these two surfaces many of the droplets were non-circular and, further, it was possible to discriminate between ice and water only at the instant of freezing. On all surfaces droplets smaller than about 10 μ also showed no change in appearance after freezing, merely giving a 'blink' at the instant of freezing. All frozen droplets that were studied melted at 0°c.

It is shown in figure 2 that despite the wide variation in distribution of freezing temperatures on the different surfaces, the lowest temperatures attained were closely the same on all surfaces. The distribution depends, too, upon the state of the surface: for example, on a silicone surface which had been used and cleaned a few times the spread of temperatures was increased towards higher temperatures (figure 2, curve *b*); a similar spread to higher temperatures, only far more pronounced, was found for a nickel surface on which sparks had been allowed to impinge. Some experiments were carried out on collodion without a silicone oil covering which made the observation of freezing easier and at the same time eliminated any possible influence of silicone oil.

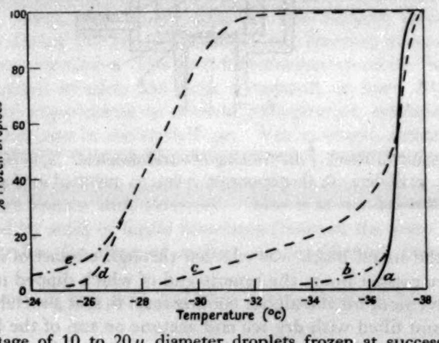


Figure 2. Percentage of 10 to 20 μ diameter droplets frozen at successively decreasing temperatures when cooled at approximately 1 deg min^{-1} on *a*, freshly prepared silicone surface; *b*, used silicone; *c*, nickel; *d*, gold; *e*, silver.

The conclusion drawn from these preliminary experiments was that while all the surfaces investigated did possess regions which stimulated freezing, the lowest freezing temperatures were independent of the nature of the surface, at least to within narrow limits. It cannot at this stage be concluded that the lowest freezing temperatures apply to homogeneous nucleation, that is to freezing which originates within the free liquid and not at an interface, for it may be that all solid surfaces influence the freezing of small droplets to the same extent or that the droplets were otherwise contaminated.

§ 4. DISTRIBUTION OF FREEZING TEMPERATURES OF INDIVIDUAL DROPLETS

The next step was to study the freezing of individual droplets, particularly those reaching the lowest temperatures. The preliminary observations showed that while no droplet froze consistently at a specific temperature, some droplets behaved more erratically than others. Further, it appeared that droplets of a given size could be classified into three groups: A, those freezing consistently within about 1°c of the lowest temperature attained; B, those having freezing temperatures spread over several degrees but occasionally reaching the lowest temperature; and C_1 those which always froze above the minimum. The relative numbers in each group depended upon the surface on which the droplets were formed.

To investigate these differences in more detail, a given field of droplets on a nickel surface was alternately frozen and melted a few hundred times and the freezing points of three selected droplets, one of each type, were noted. A cooling rate of 20 deg min^{-1} was used. The three droplets were 29, 27 and 22 μ in diameter and first froze at −36·6, −34·6 and −32°c respectively. As shown in figure 3, curve A, the lowest freezing droplet froze always within a range of 1·5°c. The next one showed a considerably wider scatter of 4°c and also a trend towards lower values. This trend is graphically illustrated by plotting cumulative frequency curves for three successive sets of 75 freezes (figure 3, curves B). The third droplet, the highest freezing one in figure 3, curve C, had freezing points distributed over 5°c with no obvious trend within the number of observations made.

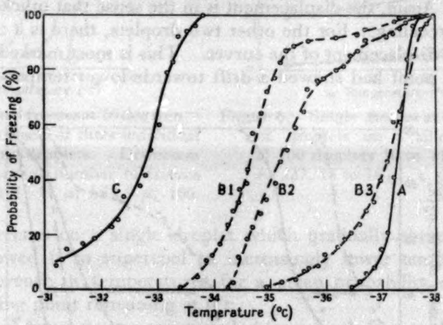


Figure 3. Cumulative frequency curves from repeated freezes at 20 deg min^{-1} of droplets A, B and C (see § 4). B_1, B_2 and B_3 are for the 1st, 2nd and 3rd sets, respectively, of 75 freezes for droplet B.

The shape of the distribution curve obtained for the lowest freezing droplet of figure 3 was found to be characteristic of all the droplets which froze consistently within about a degree of the lowest temperatures reached for a given size.

From now on these will be referred to as 'lowest freezing' or type A droplets. A large number of freezes of a single lowest freezing droplet gave the same distribution of freezing temperatures as single freezes on a large number of different droplets of the same size on a freshly prepared silicone surface (figure 2, curve *a*) showing that the majority of droplets on this surface were of type A.

§ 5. THE EFFECT OF COOLING RATE ON FREEZING POINTS

The effect of the rate of cooling on freezing points was investigated in three series of experiments. A field of ten droplets, ranging in diameter from 18 to 24μ, on a nickel surface was frozen 20 times cooling at alternately 0·2 and 2°c min^{-1}. The results for each cooling rate were consolidated into a single cumulative frequency curve. Both curves were of the same shape as figure 2, curve *c*, but the one corresponding to the higher cooling rate was depressed 0·5 ± 0·1°c in temperature throughout its length.

A similar experiment with droplets on a silicone surface cooled alternately at 1 and 10°c min^{-1} also showed a shift of the distribution curve of half a degree in the same sense.

The last fifty freezes (at 20°c min^{-1}) on the three droplets of figure 3 were alternated with freezes at 2°c min^{-1}. Fifty freezes are sufficient to determine all but the extreme ends of the distribution curves. The results are shown in figure 4. The curves for the lowest freezing droplet, A, which shrank from 27 to 22μ during these hundred freezes, are separated along their entire lengths by 0·45 ± 0·05°c. Again, the displacement is in the sense that quicker cooling leads to deeper supercooling. For the other two droplets, there is a change of shape in addition to a displacement of the curves. This is most marked for the droplet whose freezing point had showed a drift towards lower temperatures (figure 4, curves B).

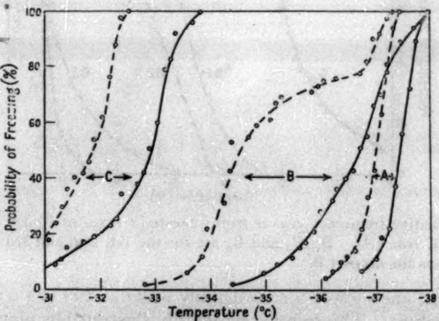


Figure 4. Repeated freezes at alternately 2 and 20 deg min^{-1} on droplets A, B and C of figure 3 (see § 4). Solid lines 20 deg min^{-1}, broken lines 2 deg min^{-1}.

Thus it is seen that a tenfold increase in the cooling rate has the effect on lowest freezing droplets of reducing the temperature of any given probability of freezing by 0·5°c. Droplets of other types behave in a more erratic and unpredictable manner. This erratic behaviour affords an explanation for the effect found by Bigg (1953 a) that a tenfold change in cooling rate caused 2°c change in freezing point, it being assumed that his droplets were contaminated (Mossop 1955, Jacobi 1955).

§ 6. The Effect of Droplet size on Freezing Points

To determine how depth of supercooling depends upon size, it is best to compare the freezing point distribution curves for droplets of different sizes. The curves in figure 5 are for individual droplets 9, 15 and 30 μ in diameter on a nickel surface. The freezing temperature of every one of these droplets reached the lowest value for droplets of its particular size and remained within 1·5°c thereof; that is, all were lowest freezing droplets. It will be seen that the curves for the smaller droplets are displaced towards lower temperatures.

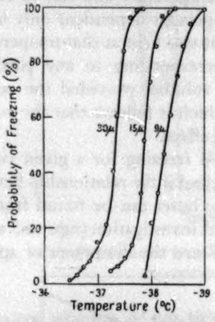


Figure 5. Repeated freezes at 20 deg min^{-1} on a nickel surface, of three individual lowest freezing droplets. Diameters and corresponding number of freezes are: 30 μ, 191; 15 μ, 65; 9 μ, 100.

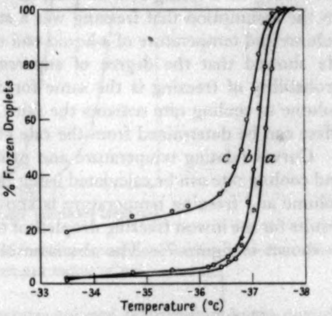


Figure 6. Single freezes at 0·5 deg min^{-1} of droplets on a silicone surface: *a*, 100 droplets 10 to 11 μ diameter; *b*, 327, 15 to 18 μ; *c*, 63, 22 to 26 μ.

Repeated freezes on a single droplet which gradually shrank in size from 30 to 20 μ showed it to supercool to increasingly lower temperatures. The maximum difference in temperatures for a given probability of freezing was 0·3°c, the melting point remaining at 0°c.

The curves of figure 6 were compounded from single freezes of a large number of droplets of three different size ranges (26 to 22, 18 to 15 and 11 to 10 μ) on a silicone surface. The lower parts of these curves are influenced by the presence of droplets which never supercool to the lowest temperatures. The percentage of these increases with mean droplet size because bigger droplets have a greater chance of coming into contact with a part of the surface that stimulates freezing. These curves show the same effect as those for the individual droplets, namely that the smaller the droplets are, the lower is the temperature for the same probability of freezing.

It is not possible to derive the precise relationship between freezing temperature and volume over the relatively restricted range of sizes investigated from the results as presented in figures 5 and 6. However, it can be seen that the reduction of supercooling following a volume increase of 50 times (diameter from 9 to 33 μ)

was just over 1°c (the curve for a 33 μ droplet was slightly to the left of the 30 μ curve shown in figure 5), a volume increase of eight times (diameter from 15 to 30 μ) was 0·7°c, and of three times (diameter from 20 to 30 μ) was 0·3°c. The curves of figure 6 point to an even smaller change of freezing temperatures with volume. It is shown in a later section that the magnitude of this effect is that for every tenfold increase in volume there is a rise of 0·5°c in the freezing temperature for any particular probability of freezing.

§ 7. DISCUSSION OF RESULTS

It is of interest to establish whether the results of the present investigation conform to a probability pattern, as would be the case if, as has been postulated, freezing is a statistical process. Bigg (1953 a) developed a relationship between the probability of freezing of a droplet, its volume, temperature and rate of cooling on the assumption that freezing was a statistical process dependent only on the volume and temperature of a liquid and the time it was kept at that temperature. He showed that the degree of supercooling corresponding to any particular probability of freezing is the same for different volumes provided the ratio of volume to cooling rate remains the same, from which it follows that the volume effect can be determined from the rate of cooling effect.

Curves relating temperature and probability of freezing for a given volume and cooling rate can be calculated using Bigg's method if the relationship between volume and freezing temperature is known. The latter can be found from the results for the lowest freezing droplets of the present investigation (figures 5 and 6) as shown in figure 7. The abscissae of figure 7 are the diameters of spheres having the same volume as the droplets which are assumed to have been hemispheres. Ordinates are temperatures for 95% probability of freezing. Temperatures for 99% are less than 0·2°c lower. While the range of diameters is too small to determine the relationship precisely, it can be seen that a straight line could be drawn to pass close to all points. The curves in figure 8 were calculated for a series of slopes of the assumed straight line. Their position relative to the temperature axis is of no significance, being determined by the particular values chosen for the volume and cooling rate. We see that an excellent match between the experimentally obtained curve for individual lowest freezing droplets (figure 5) and the calculated curves of figure 8 is in the case marked 0·5, i.e. where a tenfold

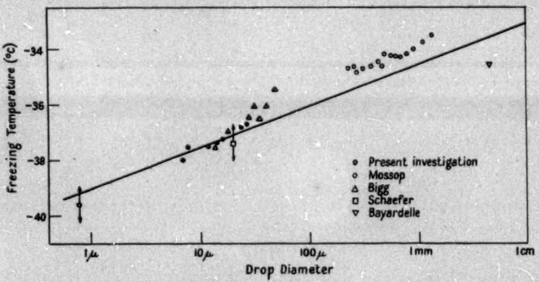


Figure 7. Temperature for 95% probability of freezing at a cooling rate of 0·5 deg min^{-1} plotted against the diameter of sphere having same volume as sample.

change in either the volume or the rate at which it is cooled corresponds to a 0·5°c change in the temperature at which it will have the same freezing probability. This agrees with the rate of cooling effect, and the volume effect (to within the limits of accuracy to which it was found), as determined experimentally. If, then, freezing temperature is directly proportional to the logarithm of the volume, the results do conform neatly to a probability pattern. Also, it follows from the above that the correction to be applied to temperatures at a cooling rate of α_1 in order that the probability of freezing be the same at α_2 is $0{\cdot}5 \log_{10}(\alpha_1/\alpha_2)$.

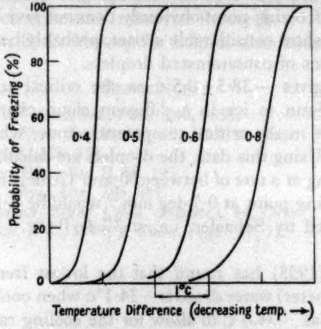


Figure 8. Calculated curves for probability of freezing : parameter is rate of cooling effect in centigrade degrees per factor of ten.

To see whether or not this temperature–volume relationship holds outside the range of sizes investigated, the line has been extrapolated in both directions (figure 7) for comparison with results of other observers. Before attempting to correlate those results which seem most likely to apply to uncontaminated samples, it is necessary to take into account differences in experimental conditions.

Bigg (1953 a, b) found the mean freezing temperature of a number of drops resting on a silicone surface or at the interface of two immiscible liquids. The results for his smallest droplets are shown in figure 7. His other results must be regarded as being for drops contaminated with freezing nuclei (Mossop 1955, Jacobi 1955). The results of Mossop (1955) shown in figure 7 are the lowest of about ten measurements on each size of the freezing point of threads of water in fine capillary tubes. Mossop also gives $-41{\cdot}2 \pm 0{\cdot}4$°c as the temperature at which 95% of the water droplets, estimated diameter $0{\cdot}8\,\mu$, froze in a small expansion cloud chamber. This value must be corrected for the fact that these rapidly growing drops would be warmer than the measured temperature of the surrounding air, and secondly, if it is to be compared with results obtained at 0·5 deg min^{-1}, for the extremely high rate of cooling. If we assume the expansion took between 0·1 and 1 second during which time the temperature fell 5°c, the above figure would have to be corrected by between 1·4 and 1·9°c if the cooling rate effect that we have found is used, i.e. to $-39{\cdot}6 \pm 0{\cdot}7$°c to correspond to a cooling rate of 0·5 deg min^{-1}. Of the results of Mossop and Bigg shown in figure 7 only the cloud chamber point represents a 95% probability of freezing; the others should be moved down about half a degree.

The results of Jacobi (1955) for $2\,\mu$ to 1 mm diameter drops, which were published while the present investigation was in progress, were also obtained by observing through a microscope the freezing of droplets condensed on a collodion surface. The drops were not covered with silicone oil, a high cooling rate (18 deg min^{-1}) had to be used and freezing points were taken to be the temperature at which the majority of droplets froze or, for the larger drops, the lowest temperatures reached by individual drops. The curve he gives lies about 2°c above the straight line proposed by Mossop. Adjustment of the points to apply to 0·5 deg min^{-1} would mean moving the curve up by almost 1°c. This method of observing a common freezing point obviously does not give lowest temperatures; moreover, the points show considerable scatter, probably because of the presence of different proportions of contaminated droplets.

Schaefer (1952) gives $-38{\cdot}5 \pm 0{\cdot}5$°c as the critical temperature at which water droplets transform to ice in a diffusion cloud chamber. The droplets fell at 1 to 2 cm sec^{-1} in the critical temperature zone, where the gradient was about 1 deg mm^{-1}. Using this data, the droplets are calculated to be about $20\,\mu$ in diameter and cooling at a rate of between 60 and 120 deg min^{-1}. Furthermore, the spontaneous freezing point at 0·5 deg min^{-1} would be $1{\cdot}1 \pm 0{\cdot}1$°c higher than the threshold observed by Schaefer, i.e. at $37{\cdot}4 \pm 0{\cdot}6$°c. This point is shown in figure 7.

Mlle Bayardelle (1955) has found that the lowest freezing temperature of 0·05 cm^3 (4·6 mm diameter) water drops is $-34{\cdot}1$°c when cooled at 0·05 deg min^{-1}. This point is plotted as $-34{\cdot}6$°c to allow for the cooling rate (figure 7).

The agreement of the cloud chamber observations with the results for the lowest freezing droplets of the present investigations suggest that the latter are for spontaneous freezing. The closeness of the results obtained by such a variety of methods to the line of figure 7, especially if allowances are made in the light of the above discussion, makes it likely that this line does represent the relationship between volume and spontaneous freezing point of water and that surface influences on lowest freezing samples are at the most of minor importance. Although in fact the relationship will probably not be linear over such a range because of the temperature variation of the physical constants of water, the important conclusion is that spontaneous freezing temperatures are little influenced by volume. Within the range of sizes of cloud droplets a million fold increase in volume will raise the temperature for a given probability of freezing by only 3°c.

§ 8. THE SURFACE FREE ENERGY AT AN ICE–WATER INTERFACE

Having concluded that the results shown in figure 7 apply to spontaneous freezing of water, it is of interest to make use of them to calculate the surface free energy σ at an ice–water interface. Following the procedure adopted by Jacobi (1955) and using the same values for physical constants, the results of the present investigation as shown in figure 7 gives values for σ somewhat higher than those found by Jacobi and which show a greater variation with temperature: e.g. when temperature $T = -38$°c, $\alpha = 16{\cdot}0$ erg cm^{-2} (Jacobi 15·5) and, over the range -33 to -40°c, $d\sigma/dT = 0{\cdot}35$ erg cm^{-2} deg^{-1} (Jacobi 0·2).

According to the theory of homogeneous nucleation as developed by Turnbull and Fisher (1949), while the nucleation rate increases continuously as the temperature is reduced, its rate of change decreases rapidly. Using $\sigma = 15{\cdot}3$ erg cm^{-2}

and taking into consideration the rate of change of σ with temperature, we find that the nucleation rate for water at -40°c increases by less than sevenfold for every degree fall in temperature. There is no sudden large increase in nucleation rate in the vicinity of the well-known transition region of water droplets near -40°c which would make this region a barrier to further supercooling. Lower temperatures could be attained with smaller droplets or more rapid cooling.

ACKNOWLEDGMENTS

This paper is published with the permission of the President of the South African Council for Scientific and Industrial Research.

I am grateful to Dr. S. C. Mossop for helpful discussions and advice.

REFERENCES

BAYARDELLE, M., 1955, *C.R. Acad. Sci., Paris*, **240**, 2553.
BIGG, E. K., 1953 a, *Proc. Phys. Soc.* B, **66**, 688; 1953 b, *Quart. J.R. Met. Soc.*, **79**, 510.
JACOBI, W., 1955, *Z. Naturf.*, **10a**, 322.
MOSSOP, S. C., 1955, *Proc. Phys. Soc.* B, **68**, 193.
MOUSSON, A., 1858, *Ann. Phys., Lpz.*, **105**, 161.
SCHAEFER, V. J., 1952, *Industr. Engng. Chem. (Industr.)*, **44**, 1300.
TURNBULL, D., and FISHER, J. C., 1949, *J. Chem. Phys.*, **17**, 71.

The Scattering of Electromagnetic Waves by Conducting Spheres and Discs

By J. S. HEY, G. S. STEWART[†], J. T. PINSON and
P. E. V. PRINCE

Radar Research Establishment, Malvern, Worcs.

MS. received 20th December 1955, and in final form 19th April 1956

Abstract. A laboratory method is described for measuring the returned signal scattered from an object in an electromagnetic field. Measurements on conducting spheres and discs are compared with theoretical formulae. A new computation of the scattering function for conducting spheres is tabulated.

§ 1. Introduction

The measurement of the scattering of electromagnetic waves from objects presents an experimental problem of some difficulty if sufficiently high accuracy is to be achieved. In order to maintain the object in specific aspects for observation some method of support is required. It is necessary to ensure that reflections from the supporting structure and from other objects in the surroundings do not produce any appreciable interfering signal at the receiver. A satisfactory laboratory method is here described in which the object is placed on a specially designed low-reflection support moving at constant speed. By utilizing the Doppler principle, reflections from the moving object can be distinguished from the surroundings. The validity of the experimental method is checked by showing that the observed variation of equivalent echoing area[‡] of conducting spheres agrees with theoretical expectations. The equivalent echoing area of other objects can then be determined, using the echoing area of a sphere as a standard. In this report the experimental method is described, and observations on conducting spheres are compared with theory. The equivalent echoing areas of conducting discs, varying in diameter from approximately 0·16 to 1·85 wavelengths are measured, and these are compared with theoretical determinations available in certain limiting conditions of disc size.

§ 2. Experimental Method

The apparatus is illustrated in figure 1. The scattering object is supported on a tall tapering mast of thin wood. Since the dielectric constant of wood is low, of the order of 2–4, it has a small reflection coefficient. The support is shaped so that any specularly reflected radiation is directed neither towards the model nor back to the receiver. The tapering of the support ensures minimum interaction with the scattering object.

[†] Now at Royal Aircraft Establishment, Farnborough, Hants.

[‡] The equivalent echoing area σ is alternatively known as the back-scattering cross section and is defined by

$$\sigma = \frac{\text{Power per unit solid angle scattered back to the transmitter}}{\text{Power flux density incident on the scattering object}}.$$

The support is carried on a wooden trolley which is moved by a constant speed motor at 16 cm sec^{-1} along a straight wooden track. The object is thus made to move along the axis of the transmitter–receiver aerial system which consists of a dipole and parabolic reflector. The distance of the object is sufficient for it to be in the Fraunhofer region of the aerial aperture. The dipole is connected to a klystron oscillator and radiates a continuous-wave signal at 3·22 cm wavelength. The signal reflected back from the model is shifted in frequency due to the Doppler effect. This reflected signal received at the aerial is combined with part of the oscillator output in a crystal mixer to give a signal at the difference frequency which is 10 c/s. After passing through a constant impedance attenuator, which has nine steps of 10 dB, the signal passes through a tuned amplifier and the output is measured on a valve voltmeter. The amplifier input–output characteristic was calibrated by means of a signal generator.

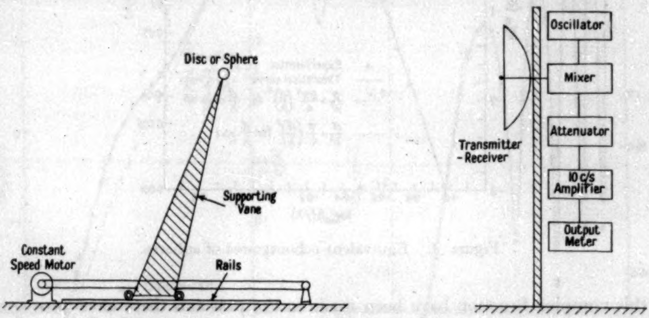


Figure 1. Equipment for measuring echoing areas.

To ensure that the incident field presented a plane wave of constant amplitude in the vicinity of the object, this region was explored with the aid of a small metal prism rotating so that successive faces became normal to the incident beam at a frequency of 10 c/s. The receiver response then depends on both the amplitude and the phase of the incident radiation on the prism and it was possible to plot phase fronts and the amplitude distributions over them.

The field intensity at the trolley was sufficiently low for its reflected signal to be negligible. All surrounding objects were at distances of more than 50 λ. Reflections from the walls should not in general contribute to the Doppler signal frequency, but to eliminate any possible effects, including those of vibration, radio-absorbent material was interposed. This also ensured that any disturbance of the incident field on the scattering object by reflection from the walls was of negligible intensity.

§ 3. SPHERES

The measurements on conducting spheres were taken principally to check the validity of the experimental method and to provide a standard for determining the echoing areas of other objects. The experimental values are the points plotted in figure 2 for spheres of diameter ranging from about 0·17 λ to 2·0 λ.

The experimental results are relative and have been scaled in figure 2 by equalizing the echoing area at an arbitrarily chosen point to the theoretically derived value computed from Mie's (1908) formula. The theoretical function defining the equivalent echoing area of a sphere is given in Appendix 1. Various computations of this complex function have been made in the past and there have been small but noticeable differences between them. In an attempt to remove the discrepancies a new series of values have recently been computed from Mie's formula by the Scientific Computing Service Ltd., and these are tabulated in Appendix 1. The full-line curve in figure 2 has been obtained from this theoretical computation. It is seen in figure 2 that there is a very close agreement between the experimental results and theoretical values. The supporting vane did not give a signal greater than $0{\cdot}001\,\lambda^2$ which was the minimum detectable signal.

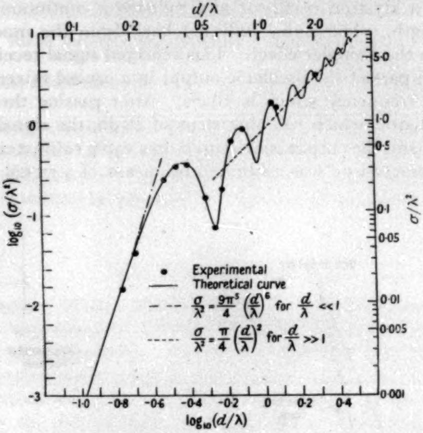


Figure 2. Equivalent echoing area of spheres.

When the diameter *d* of the sphere is very small compared with the wavelength the equivalent echoing area σ approaches $9\pi^5 d^6/4\lambda^4$, whilst for large spheres reflection is specular and the echoing area approaches the cross-sectional area, so that $\sigma = \frac{1}{4}\pi d^2$. These limiting values are plotted as broken lines in figure 2.

§ 4. DISCS

4.1. *Discs Normal to the Direction of Incidence*

A plane surface of dimensions large compared with the wavelength has an equivalent echoing area σ at normal incidence given by $\sigma = 4\pi S^2/\lambda^2$ where S is the surface area. If the plane surface is a disc of radius d, then $\sigma = \pi^3 d^4/4\lambda^2$.

The derivation of the formula (see, for example, Kerr 1951, p. 457) assumes that the current distribution is the same as in an infinite plane surface. It is consequently of interest to ascertain experimentally for what disc dimensions the approximation is valid. The experimental results in figure 3 indicate that for d greater than $1{\cdot}5\,\lambda$ the discrepancy becomes insignificant.

For a small disc, the echoing area at normal incidence may be deduced theoretically from the electric dipole moment induced by the electric vector of the incident wave. The echoing area σ (see Appendix 2) is given by $\sigma = 16\pi^3 d^6/9\lambda^4$.

The experimental results in figure 3 suggest that although these may be about 10% greater than the theoretical formula for the smallest measured disc $d = 0{\cdot}16\,\lambda$, the experimental and theoretical curves are converging as the disc size diminishes. The experimental measurements become difficult when the discs are so small. The results of two sets of measurements shown in figure 3 indicate the experimental scatter, which for the smallest disc is of the same order as the differences between the experimental and theoretical results. The broken line in figure 3 joins the theoretical results for large and small discs. The experimental results show a divergence in the region of $d \simeq \lambda/2$, indicating a partial resonance.

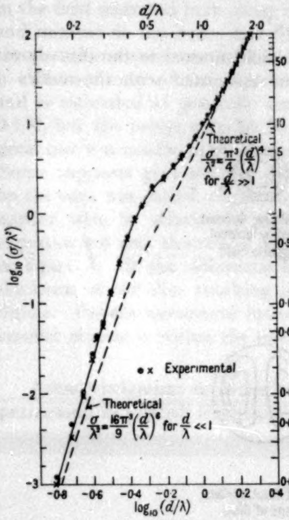


Figure 3. Equivalent echoing area of thin discs normal to the direction of incidence.

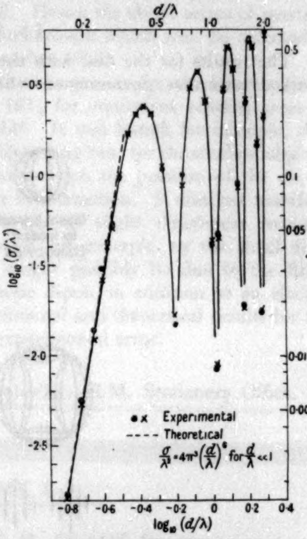


Figure 4. Equivalent echoing area of thin discs normal to the magnetic vector.

4.2. *Discs Normal to the Magnetic Vector*

If a disc is in the plane containing the direction of incidence and the electric vector, then re-radiated fields occur as a consequence not only of the currents induced by the electric field vector tangential to the disc but also due to the currents induced by the magnetic field vector normal to the disc. For large discs, the phase difference over the disc introduces further complexity and no theoretical evaluations are available. The experimental results, illustrated in figure 4, show a pronounced resonance phenomenon, which is discussed in more detail later (§ 5).

The theoretical derivation of the formula for small discs is given in Appendix 2. The scattered field is due to the combined effects of the induced electric and magnetic dipoles, and the equivalent echoing area σ is greater by a factor 9/4 than for the disc normal to the direction of incidence. Thus $\sigma = 4\pi^3 d^6/\lambda^4$, and the experimental results plotted in figure 4 show that for small discs there is close agreement with theoretical expectations.

§ 5. DISCUSSION

The results for the disc with the magnetic field normal to the disc show a marked resonance phenomenon which may be associated with the modes of oscillation. The following qualitative explanation is suggested. As the diameter is increased up to $\lambda/2$, the total electric dipole moment will increase in magnitude without any phase reversal. Similarly the total magnetic flux normal to the disc will increase and produce a greater circulating current. The resultant electric and magnetic dipole moments produce additive fields scattered back in the incident direction and a maximum may be expected as the diameter approaches $\lambda/2$. This oscillation is illustrated in figure 5 (*a*).

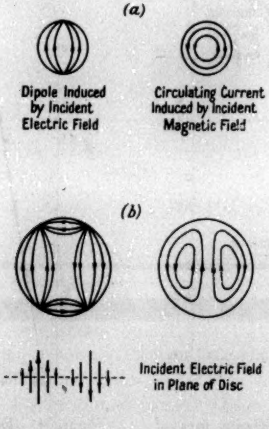

(a)
Dipole Induced by Incident Electric Field
Circulating Current Induced by Incident Magnetic Field
(b)
Incident Electric Field in Plane of Disc

Figure 5. Modes of oscillation of discs normal to the magnetic vector.

As the diameter is increased so that a phase reversal occurs across the disc, a second mode of oscillation will develop as illustrated in figure 5 (*b*) with two opposite electric dipoles and two opposite systems of circulating currents. The radiation from the opposing electric dipoles and magnetic dipoles, separated by $\lambda/2$ in the plane of the disc, reinforce back in the incident direction. This radiated field is however opposite in phase to that due to the first mode of oscillation. At an appropriate disc size the two fields are equal in magnitude and thus cancel out. As the diameter is further increased, the second mode of oscillation illustrated in figure 5 (*b*) becomes the dominant mode and a second maximum occurs. With further increase of diameter, other modes develop and the process of successive deep minima and maxima in the reflected amplitude continues.

The three-dimensional oscillatory system for a sphere in an electromagnetic field is more complex but there are features of general similarity. As may be expected, the echoing area is greater but with less pronounced maxima and minima.

For a disc normal to the direction of incidence there is no phase difference in the field scattered back along the normal. Hence the above series of maxima and minima do not occur and the only enhancement which may be detected is the partial resonance as the diameter approaches $\lambda/2$.

The accuracy of the experimental method is greater for the larger objects and is estimated as generally better than 10% for equivalent echoing areas of $0{\cdot}1\lambda^2$, but the errors may be 20% at $0{\cdot}01\lambda^2$. It was found, for example, that great care was needed in the design of the supporting vane for the smallest objects. Some supports gave slightly different results when the position of the object on the vane was varied, suggesting evidence of interaction. It does not therefore appear wise to attach too much significance to slight differences between experimental and theoretical values as shown, for example, by the small discs in figure 3. If the difference is real, it might possibly be due to the finite thickness of the disc resulting in a magnetic dipole in addition to an electric dipole. But the agreement between experimental and theoretical results for the smaller objects is within the tolerance of experimental error.

Acknowledgment is made to the Controller, H.M. Stationery Office, for permission to publish this paper.

APPENDIX 1

The authors are indebted to Mr. P. H. Blundell for arranging a new computation by the Scientific Computing Service Limited of the scattering function for a perfectly conducting sphere derived by Mie (1908). The results of this computation are tabulated below. The values of the function f are given in terms of $x = 2\pi a/\lambda$ where a is the radius and f is defined by

$$f = \left| \sum_{n=1}^{\infty} (-1)^n (2n+1) \left[\frac{\mathrm{j}_n(x)}{\mathrm{h}_n^{(1)}(x)} - \frac{(d/dx)\{x\mathrm{j}_n(x)\}}{(d/dx)\{x\mathrm{h}_n^{(1)}(x)\}} \right] \right|,$$

where $\mathrm{j}_n(x)$ and $\mathrm{h}_n^{(1)}(x)$ are spherical Bessel and Hankel functions related to the cylindrical functions by

$$\mathrm{j}_n(x) = (\pi/2x)^{1/2}\, \mathrm{J}_{n+\frac{1}{2}}(x),$$
$$\mathrm{h}_n^{(1)}(x) = (\pi/2x)^{1/2}\, \mathrm{H}_{n+\frac{1}{2}}^{(1)}(x)$$

where $\mathrm{H}^{(1)}$ is a Hankel function of the first kind with time convention $\exp(-i\omega t)$.

The equivalent echoing area (σ) of a sphere is given by $\sigma = \lambda^2 f^2/4\pi$.

x	*f*	*x*	*f*	*x*	*f*	*x*	*f*	*x*	*f*
0·00	0·0000	0·55	0·2396	1·10	1·0359	1·65	0·5271	2·20	1·4702
0·01	0·0000	0·56	0·2523	1·11	1·0408	1·66	0·5138	2·21	1·4874
0·02	0·0000	0·57	0·2653	1·12	1·0451	1·67	0·5017	2·22	1·5039
0·03	0·0000	0·58	0·2788	1·13	1·0488	1·68	0·4910	2·23	1·5197
0·04	0·0001	0·59	0·2925	1·14	1·0518	1·69	0·4819	2·24	1·5347
0·05	0·0002	0·60	0·3067	1·15	1·0542	1·70	0·4744	2·25	1·5490
0·06	0·0003	0·61	0·3212	1·16	1·0561	1·71	0·4689	2·26	1·5625
0·07	0·0005	0·62	0·3360	1·17	1·0573	1·72	0·4653	2·27	1·5753
0·08	0·0008	0·63	0·3512	1·18	1·0578	1·73	0·4639	2·28	1·5872
0·09	0·0011	0·64	0·3667	1·19	1·0578	1·74	0·4646	2·29	1·5984
0·10	0·0015	0·65	0·3825	1·20	1·0572	1·75	0·4677	2·30	1·6087
0·11	0·0020	0·66	0·3986	1·21	1·0560	1·76	0·4730	2·31	1·6183
0·12	0·0026	0·67	0·4150	1·22	1·0541	1·77	0·4807	2·32	1·6270
0·13	0·0033	0·68	0·4317	1·23	1·0517	1·78	0·4905	2·33	1·6349
0·14	0·0041	0·69	0·4486	1·24	1·0487	1·79	0·5024	2·34	1·6419
0·15	0·0051	0·70	0·4657	1·25	1·0451	1·80	0·5162	2·35	1·6481
0·16	0·0061	0·71	0·4831	1·26	1·0409	1·81	0·5319	2·36	1·6535
0·17	0·0074	0·72	0·5006	1·27	1·0361	1·82	0·5493	2·37	1·6580
0·18	0·0087	0·73	0·5183	1·28	1·0308	1·83	0·5683	2·38	1·6617
0·19	0·0103	0·74	0·5361	1·29	1·0248	1·84	0·5886	2·39	1·6645
0·20	0·0120	0·75	0·5541	1·30	1·0183	1·85	0·6102	2·40	1·6665
0·21	0·0138	0·76	0·5721	1·31	1·0112	1·86	0·6329	2·41	1·6676
0·22	0·0159	0·77	0·5902	1·32	1·0036	1·87	0·6565	2·42	1·6678
0·23	0·0182	0·78	0·6083	1·33	0·9954	1·88	0·6810	2·43	1·6672
0·24	0·0206	0·79	0·6264	1·34	0·9866	1·89	0·7061	2·44	1·6658
0·25	0·0233	0·80	0·6445	1·35	0·9773	1·90	0·7319	2·45	1·6635
0·26	0·0262	0·81	0·6625	1·36	0·9675	1·91	0·7582	2·46	1·6603
0·27	0·0293	0·82	0·6804	1·37	0·9571	1·92	0·7849	2·47	1·6563
0·28	0·0327	0·83	0·6982	1·38	0·9462	1·93	0·8119	2·48	1·6515
0·29	0·0363	0·84	0·7159	1·39	0·9347	1·94	0·8391	2·49	1·6459
0·30	0·0402	0·85	0·7333	1·40	0·9228	1·95	0·8666	2·50	1·6394
0·31	0·0443	0·86	0·7506	1·41	0·9103	1·96	0·8941	2·51	1·6321
0·32	0·0487	0·87	0·7675	1·42	0·8974	1·97	0·9217	2·52	1·6240
0·33	0·0533	0·88	0·7842	1·43	0·8839	1·98	0·9492	2·53	1·6151
0·34	0·0583	0·89	0·8006	1·44	0·8700	1·99	0·9767	2·54	1·6055
0·35	0·0635	0·90	0·8167	1·45	0·8557	2·00	1·0041	2·55	1·5951
0·36	0·0691	0·91	0·8324	1·46	0·8409	2·01	1·0313	2·56	1·5839
0·37	0·0749	0·92	0·8477	1·47	0·8257	2·02	1·0582	2·57	1·5721
0·38	0·0811	0·93	0·8626	1·48	0·8101	2·03	1·0850	2·58	1·5595
0·39	0·0876	0·94	0·8771	1·49	0·7941	2·04	1·1114	2·59	1·5463
0·40	0·0944	0·95	0·8911	1·50	0·7778	2·05	1·1375	2·60	1·5324
0·41	0·1016	0·96	0·9046	1·51	0·7612	2·06	1·1633	2·61	1·5179
0·42	0·1091	0·97	0·9177	1·52	0·7443	2·07	1·1886	2·62	1·5028
0·43	0·1169	0·98	0·9302	1·53	0·7272	2·08	1·2135	2·63	1·4872
0·44	0·1251	0·99	0·9422	1·54	0·7099	2·09	1·2380	2·64	1·4710
0·45	0·1337	1·00	0·9536	1·55	0·6925	2·10	1·2620	2·65	1·4543
0·46	0·1426	1·01	0·9645	1·56	0·6749	2·11	1·2854	2·66	1·4372
0·47	0·1519	1·02	0·9748	1·57	0·6574	2·12	1·3084	2·67	1·4197
0·48	0·1615	1·03	0·9846	1·58	0·6399	2·13	1·3308	2·68	1·4018
0·49	0·1715	1·04	0·9937	1·59	0·6226	2·14	1·3526	2·69	1·3837
0·50	0·1819	1·05	1·0023	1·60	0·6054	2·15	1·3738	2·70	1·3653
0·51	0·1927	1·06	1·0102	1·61	0·5886	2·16	1·3944	2·71	1·3467
0·52	0·2039	1·07	1·0176	1·62	0·5722	2·17	1·4143	2·72	1·3279
0·53	0·2154	1·08	1·0243	1·63	0·5564	2·18	1·4336	2·73	1·3091
0·54	0·2274	1·09	1·0304	1·64	0·5413	2·19	1·4522	2·74	1·2903

x	*f*	*x*	*f*	*x*	*f*	*x*	*f*	*x*	*f*
2·75	1·2716	3·30	1·7993	3·85	2·0368	4·40	2·0944	4·95	2·7680
2·76	1·2530	3·31	1·8251	3·86	2·0194	4·41	2·1218	4·96	2·7563
2·77	1·2346	3·32	1·8504	3·87	2·0017	4·42	2·1493	4·97	2·7439
2·78	1·2166	3·33	1·8751	3·88	1·9838	4·43	2·1770	4·98	2·7308
2·79	1·1990	3·34	1·8994	3·89	1·9657	4·44	2·2046	4·99	2·7171
2·80	1·1818	3·35	1·9230	3·90	1·9474	4·45	2·2323	5·00	2·7028
2·81	1·1653	3·36	1·9461	3·91	1·9291	4·46	2·2598	5·01	2·6880
2·82	1·1494	3·37	1·9685	3·92	1·9108	4·47	2·2871	5·02	2·6727
2·83	1·1343	3·38	1·9903	3·93	1·8925	4·48	2·3143	5·03	2·6570
2·84	1·1201	3·39	2·0114	3·94	1·8744	4·49	2·3412	5·04	2·6408
2·85	1·1069	3·40	2·0317	3·95	1·8564	4·50	2·3678	5·05	2·6243
2·86	1·0948	3·41	2·0514	3·96	1·8388	4·51	2·3940	5·06	2·6075
2·87	1·0839	3·42	2·0702	3·97	1·8215	4·52	2·4198	5·07	2·5904
2·88	1·0742	3·43	2·0883	3·98	1·8046	4·53	2·4452	5·08	2·5732
2·89	1·0660	3·44	2·1055	3·99	1·7882	4·54	2·4700	5·09	2·5558
2·90	1·0592	3·45	2·1220	4·00	1·7724	4·55	2·4944	5·10	2·5384
2·91	1·0539	3·46	2·1375	4·01	1·7572	4·56	2·5181	5·11	2·5209
2·92	1·0502	3·47	2·1523	4·02	1·7428	4·57	2·5413	5·12	2·5035
2·93	1·0482	3·48	2·1661	4·03	1·7292	4·58	2·5637	5·13	2·4862
2·94	1·0479	3·49	2·1790	4·04	1·7165	4·59	2·5855	5·14	2·4691
2·95	1·0493	3·50	2·1910	4·05	1·7047	4·60	2·6066	5·15	2·4522
2·96	1·0524	3·51	2·2021	4·06	1·6940	4·61	2·6269	5·16	2·4356
2·97	1·0573	3·52	2·2123	4·07	1·6844	4·62	2·6465	5·17	2·4194
2·98	1·0640	3·53	2·2215	4·08	1·6760	4·63	2·6653	5·18	2·4037
2·99	1·0724	3·54	2·2298	4·09	1·6688	4·64	2·6832	5·19	2·3884
3·00	1·0825	3·55	2·2371	4·10	1·6629	4·65	2·7002	5·20	2·3738
3·01	1·0942	3·56	2·2434	4·11	1·6583	4·66	2·7164	5·21	2·3598
3·02	1·1075	3·57	2·2488	4·12	1·6551	4·67	2·7317	5·22	2·3465
3·03	1·1223	3·58	2·2532	4·13	1·6534	4·68	2·7461	5·23	2·3340
3·04	1·1386	3·59	2·2566	4·14	1·6531	4·69	2·7596	5·24	2·3223
3·05	1·1562	3·60	2·2591	4·15	1·6543	4·70	2·7721	5·25	2·3116
3·06	1·1751	3·61	2·2606	4·16	1·6570	4·71	2·7836	5·26	2·3018
3·07	1·1952	3·62	2·2611	4·17	1·6613	4·72	2·7942	5·27	2·2931
3·08	1·2164	3·63	2·2606	4·18	1·6670	4·73	2·8038	5·28	2·2854
3·09	1·2386	3·64	2·2592	4·19	1·6742	4·74	2·8124	5·29	2·2789
3·10	1·2618	3·65	2·2568	4·20	1·6829	4·75	2·8200	5·30	2·2736
3·11	1·2858	3·66	2·2535	4·21	1·6931	4·76	2·8266	5·31	2·2695
3·12	1·3106	3·67	2·2492	4·22	1·7047	4·77	2·8322	5·32	2·2666
3·13	1·3360	3·68	2·2441	4·23	1·7177	4·78	2·8367	5·33	2·2651
3·14	1·3620	3·69	2·2380	4·24	1·7320	4·79	2·8403	5·34	2·2649
3·15	1·3884	3·70	2·2310	4·25	1·7477	4·80	2·8429	5·35	2·2660
3·16	1·4154	3·71	2·2232	4·26	1·7646	4·81	2·8445	5·36	2·2685
3·17	1·4426	3·72	2·2145	4·27	1·7826	4·82	2·8451	5·37	2·2724
3·18	1·4702	3·73	2·2050	4·28	1·8018	4·83	2·8447	5·38	2·2776
3·19	1·4979	3·74	2·1947	4·29	1·8221	4·84	2·8433	5·39	2·2842
3·20	1·5258	3·75	2·1836	4·30	1·8433	4·85	2·8409	5·40	2·2922
3·21	1·5537	3·76	2·1717	4·31	1·8655	4·86	2·8376	5·41	2·3015
3·22	1·5817	3·77	2·1591	4·32	1·8885	4·87	2·8334	5·42	2·3122
3·23	1·6096	3·78	2·1459	4·33	1·9123	4·88	2·8282	5·43	2·3241
3·24	1·6374	3·79	2·1319	4·34	1·9367	4·89	2·8222	5·44	2·3373
3·25	1·6651	3·80	2·1174	4·35	1·9619	4·90	2·8152	5·45	2·3517
3·26	1·6925	3·81	2·1023	4·36	1·9875	4·91	2·8074	5·46	2·3674
3·27	1·7197	3·82	2·0866	4·37	2·0137	4·92	2·7988	5·47	2·3841
3·28	1·7466	3·83	2·0704	4·38	2·0403	4·93	2·7893	5·48	2·4020
3·29	1·7732	3·84	2·0538	4·39	2·0672	4·94	2·7790	5·49	2·4209

x	f	x	f	x	f	x	f	x	f
5·50	2·4408	6·05	3·4235	6·60	2·9057	7·15	3·9792	7·70	3·5049
5·51	2·4616	6·06	3·4205	6·61	2·9143	7·16	3·9863	7·71	3·5014
5·52	2·4832	6·07	3·4167	6·62	2·9243	7·17	3·9925	7·72	3·4990
5·53	2·5057	6·08	3·4119	6·63	2·9354	7·18	3·9976	7·73	3·4977
5·54	2·5289	6·09	3·4062	6·64	2·9477	7·19	4·0018	7·74	3·4975
5·55	2·5528	6·10	3·3996	6·65	2·9612	7·20	4·0051	7·75	3·4985
5·56	2·5772	6·11	3·3922	6·66	2·9758	7·21	4·0073	7·76	3·5007
5·57	2·6022	6·12	3·3840	6·67	2·9915	7·22	4·0086	7·77	3·5040
5·58	2·6277	6·13	3·3750	6·68	3·0082	7·23	4·0089	7·78	3·5085
5·59	2·6536	6·14	3·3652	6·69	3·0259	7·24	4·0083	7·79	3·5142
5·60	2·6798	6·15	3·3547	6·70	3·0446	7·25	4·0068	7·80	3·5211
5·61	2·7063	6·16	3·3435	6·71	3·0641	7·26	4·0043	7·81	3·5292
5·62	2·7330	6·17	3·3316	6·72	3·0846	7·27	4·0009	7·82	3·5384
5·63	2·7599	6·18	3·3191	6·73	3·1058	7·28	3·9966	7·83	3·5487
5·64	2·7868	6·19	3·3060	6·74	3·1277	7·29	3·9915	7·84	3·5601
5·65	2·8138	6·20	3·2923	6·75	3·1503	7·30	3·9855	7·85	3·5727
5·66	2·8407	6·21	3·2782	6·76	3·1735	7·31	3·9787	7·86	3·5863
5·67	2·8675	6·22	3·2635	6·77	3·1973	7·32	3·9710	7·87	3·6009
5·68	2·8942	6·23	3·2485	6·78	3·2216	7·33	3·9627	7·88	3·6165
5·69	2·9207	6·24	3·2331	6·79	3·2463	7·34	3·9535	7·89	3·6331
5·70	2·9469	6·25	3·2173	6·80	3·2714	7·35	3·9437	7·90	3·6506
5·71	2·9728	6·26	3·2013	6·81	3·2968	7·36	3·9331	7·91	3·6690
5·72	2·9983	6·27	3·1851	6·82	3·3224	7·37	3·9220	7·92	3·6881
5·73	3·0234	6·28	3·1687	6·83	3·3482	7·38	3·9102	7·93	3·7081
5·74	3·0481	6·29	3·1522	6·84	3·3742	7·39	3·8978	7·94	3·7288
5·75	3·0723	6·30	3·1357	6·85	3·4002	7·40	3·8849	7·95	3·7501
5·76	3·0959	6·31	3·1191	6·86	3·4262	7·41	3·8715	7·96	3·7721
5·77	3·1190	6·32	3·1027	6·87	3·4522	7·42	3·8577	7·97	3·7947
5·78	3·1414	6·33	3·0863	6·88	3·4781	7·43	3·8435	7·98	3·8177
5·79	3·1632	6·34	3·0702	6·89	3·5038	7·44	3·8289	7·99	3·8412
5·80	3·1843	6·35	3·0543	6·90	3·5294	7·45	3·8141	8·00	3·8651
5·81	3·2047	6·36	3·0387	6·91	3·5546	7·46	3·7989	8·01	3·8893
5·82	3·2243	6·37	3·0235	6·92	3·5796	7·47	3·7836	8·02	3·9138
5·83	3·2431	6·38	3·0087	6·93	3·6042	7·48	3·7681	8·03	3·9385
5·84	3·2611	6·39	2·9945	6·94	3·6284	7·49	3·7526	8·04	3·9634
5·85	3·2783	6·40	2·9808	6·95	3·6521	7·50	3·7370	8·05	3·9885
5·86	3·2946	6·41	2·9677	6·96	3·6754	7·51	3·7214	8·06	4·0135
5·87	3·3101	6·42	2·9553	6·97	3·6981	7·52	3·7059	8·07	4·0386
5·88	3·3246	6·43	2·9437	6·98	3·7202	7·53	3·6905	8·08	4·0636
5·89	3·3382	6·44	2·9329	6·99	3·7417	7·54	3·6753	8·09	4·0885
5·90	3·3509	6·45	2·9229	7·00	3·7626	7·55	3·6604	8·10	4·1132
5·91	3·3626	6·46	2·9139	7·01	3·7828	7·56	3·6457	8·11	4·1378
5·92	3·3734	6·47	2·9059	7·02	3·8023	7·57	3·6315	8·12	4·1620
5·93	3·3832	6·48	2·8988	7·03	3·8210	7·58	3·6176	8·13	4·1860
5·94	3·3920	6·49	2·8928	7·04	3·8390	7·59	3·6043	8·14	4·2096
5·95	3·3998	6·50	2·8879	7·05	3·8562	7·60	3·5915	8·15	4·2329
5·96	3·4066	6·51	2·8842	7·06	3·8725	7·61	3·5792	8·16	4·2556
5·97	3·4125	6·52	2·8816	7·07	3·8880	7·62	3·5677	8·17	4·2779
5·98	3·4173	6·53	2·8803	7·08	3·9026	7·63	3·5568	8·18	4·2997
5·99	3·4211	6·54	2·8801	7·09	3·9163	7·64	3·5468	8·19	4·3209
6·00	3·4240	6·55	2·8812	7·10	3·9291	7·65	3·5375	8·20	4·3415
6·01	3·4258	6·56	2·8836	7·11	3·9410	7·66	3·5291	8·21	4·3615
6·02	3·4267	6·57	2·8872	7·12	3·9520	7·67	3·5216	8·22	4·3808
6·03	3·4266	6·58	2·8921	7·13	3·9620	7·68	3·5150	8·23	4·3993
6·04	3·4255	6·59	2·8982	7·14	3·9711	7·69	3·5094	8·24	4·4172

x	f	x	f	x	f	x	f	x	f
8·25	4·4343	8·60	4·4799	8·95	4·1167	9·30	4·6980	9·65	5·1756
8·26	4·4506	8·61	4·4674	8·96	4·1186	9·31	4·7218	9·66	5·1743
8·27	4·4661	8·62	4·4545	8·97	4·1216	9·32	4·7454	9·67	5·1722
8·28	4·4808	8·63	4·4412	8·98	4·1256	9·33	4·7687	9·68	5·1693
8·29	4·4947	8·64	4·4275	8·99	4·1308	9·34	4·7917	9·69	5·1655
8·30	4·5076	8·65	4·4136	9·00	4·1371	9·35	4·8143	9·70	5·1609
8·31	4·5197	8·66	4·3994	9·01	4·1445	9·36	4·8366	9·71	5·1556
8·32	4·5308	8·67	4·3851	9·02	4·1529	9·37	4·8584	9·72	5·1495
8·33	4·5411	8·68	4·3706	9·03	4·1624	9·38	4·8798	9·73	5·1426
8·34	4·5504	8·69	4·3560	9·04	4·1730	9·39	4·9006	9·74	5·1351
8·35	4·5588	8·70	4·3413	9·05	4·1846	9·40	4·9209	9·75	5·1269
8·36	4·5663	8·71	4·3267	9·06	4·1972	9·41	4·9406	9·76	5·1180
8·37	4·5728	8·72	4·3121	9·07	4·2108	9·42	4·9597	9·77	5·1086
8·38	4·5784	8·73	4·2977	9·08	4·2254	9·43	4·9781	9·78	5·0985
8·39	4·5829	8·74	4·2835	9·09	4·2409	9·44	4·9958	9·79	5·0879
8·40	4·5866	8·75	4·2695	9·10	4·2572	9·45	5·0128	9·80	5·0769
8·41	4·5893	8·76	4·2557	9·11	4·2744	9·46	5·0291	9·81	5·0653
8·42	4·5911	8·77	4·2423	9·12	4·2924	9·47	5·0446	9·82	5·0534
8·43	4·5919	8·78	4·2294	9·13	4·3112	9·48	5·0593	9·83	5·0410
8·44	4·5918	8·79	4·2168	9·14	4·3306	9·49	5·0732	9·84	5·0284
8·45	4·5908	8·80	4·2048	9·15	4·3508	9·50	5·0863	9·85	5·0154
8·46	4·5889	8·81	4·1934	9·16	4·3715	9·51	5·0986	9·86	5·0022
8·47	4·5861	8·82	4·1825	9·17	4·3929	9·52	5·1099	9·87	4·9888
8·48	4·5825	8·83	4·1723	9·18	4·4147	9·53	5·1204	9·88	4·9753
8·49	4·5780	8·84	4·1628	9·19	4·4370	9·54	5·1300	9·89	4·9616
8·50	4·5726	8·85	4·1541	9·20	4·4597	9·55	5·1387	9·90	4·9480
8·51	4·5665	8·86	4·1462	9·21	4·4828	9·56	5·1465	9·91	4·9343
8·52	4·5596	8·87	4·1391	9·22	4·5062	9·57	5·1534	9·92	4·9207
8·53	4·5519	8·88	4·1329	9·23	4·5299	9·58	5·1593	9·93	4·9071
8·54	4·5436	8·89	4·1276	9·24	4·5537	9·59	5·1644	9·94	4·8938
8·55	4·5345	8·90	4·1232	9·25	4·5777	9·60	5·1685	9·95	4·8806
8·56	4·5247	8·91	4·1199	9·26	4·6018	9·61	5·1717	9·96	4·8677
8·57	4·5144	8·92	4·1175	9·27	4·6259	9·62	5·1741	9·97	4·8551
8·58	4·5034	8·93	4·1162	9·28	4·6500	9·63	5·1755	9·98	4·8429
8·59	4·4919	8·94	4·1159	9·29	4·6741	9·64	5·1760	9·99	4·8311
								10·00	4·8198

APPENDIX 2

The following is a simplified statement to show how the equivalent echoing area of a small conducting disc may be derived. The method mainly applies an approach used by Bethe (1944) in solving diffraction at a small aperture to the problem of scattering by a disc in a way which emphasizes the physical interpretation.

(*a*) A plane transverse electromagnetic wave incident normally on a small disc produces an oscillating electric dipole moment excited by the tangential electric field vector. A thin circular disc approximates to a thin oblate spheroid, and the electrostatic problem of determining the potential due to the induced charges when the oblate spheroid is placed in a uniform parallel field is solved in several textbooks (see, for example, Stratton 1941, p. 209). The potential at distance r is

$$\frac{d^3E}{6\pi}\frac{\cos\theta}{r^2} \quad \text{(m.k.s. units)}$$

where θ is the angle between the field $\mathbf{E}$ and the radius vector $\mathbf{r}$ and d is the diameter of the disc.

This is equal to the potential due to an electric dipole of moment

$$\mathbf{p} = \tfrac{2}{3}\epsilon_0 d^3 \mathbf{E} \quad (\epsilon_0 \text{ is the permittivity of free space}).$$

If the electric field intensity in the incident wave is $\mathbf{E} = \mathbf{E}_0 \sin \omega t$, then since the dimensions are small and phase retardation is negligible, the re-radiated field is equal to that from a Hertzian dipole of moment $\mathbf{p} = \mathbf{p}_0 \sin \omega t$.

The magnitude of the radiation field scattered back in the incident direction is therefore

$$E_s = \frac{\pi}{\epsilon_0 \lambda^2 r}\, p = \frac{2\pi d^3}{3\lambda^2 r} E.$$

By definition, the equivalent echoing area is

$$\sigma = 4\pi r^2 \frac{|E_s|^2}{|E|^2} = \frac{16}{9}\frac{\pi^3 d^6}{\lambda^4}.$$

(*b*) If the magnetic vector is normal to the disc, there is not only an electric dipole $\mathbf{p}$ parallel to the electric field, but also a magnetic dipole due to the circulating currents induced by the magnetic field. These currents may be calculated as follows:

As the disc is a perfect conductor, the electromagnetic field in it is zero. Since the incident magnetic field H is normal to the disc a current distribution is required which gives a normal field in the disc equal and opposite to the applied field H. An additional current distribution is required which does not affect the charge distributions induced by the electric field. The current distribution must consequently consist of closed lines of circulating current. This problem may be solved as follows by introducing the vector potential $\mathbf{A}$ defined by $\mathbf{H} = \operatorname{curl} \mathbf{A}$. If the x axis is along H, then the circular current distribution must produce at the disc a field $H_x = -H$, $H_y = 0$, $H_z = 0$. Hence the components of vector potential at the disc are $A_x = 0$, $A_y = \tfrac{1}{2}zH$, $A_z = -\tfrac{1}{2}yH$.

For a small disc the phase retardation may be neglected and the relation between the vector potential $\mathbf{A}$ and the surface current density over the elements of area dS of the disc is

$$\mathbf{A}(r) = \frac{1}{4\pi}\int_S \frac{\mathbf{K}(r')}{|\mathbf{r} - \mathbf{r}'|}\, dS$$

where $\mathbf{r}$ and $\mathbf{r}'$ are the radius vectors to the field and source points respectively.

Hence the current distribution on the disc is given by

$$\frac{1}{4\pi}\int_S \frac{K_y'\, dS}{|\mathbf{r} - \mathbf{r}'|} = +\tfrac{1}{2}zH \quad \text{and} \quad \frac{1}{4\pi}\int_S \frac{K_z'\, dS}{|\mathbf{r} - \mathbf{r}'|} = -\tfrac{1}{2}yH.$$

The required solutions to these equations are:

$$K_y' = \frac{4}{\pi}\frac{1}{(a^2 - r'^2)^{1/2}} z'H \quad \text{and} \quad K_z' = -\frac{4}{\pi}\frac{1}{(a^2 - r'^2)^{1/2}} y'H$$

so that

$$\mathbf{K} = \frac{4}{\pi}\frac{1}{(a^2 - r'^2)^{1/2}} \mathbf{r}' \times \mathbf{H}$$

which is a circular current with amplitude proportional to $r'/(a^2 - r'^2)^{1/2}$.

The magnetic dipole moment associated with a circular current I of radius r has magnitude $\mu_0 I \pi r^2$, where μ_0 is the permeability of free space.

Hence the magnetic dipole moment due to the current density K is

$$M=\mu_0\int_0^a K\,dr'\pi r'^2$$

$$=4\mu_0 H\int_0^a \frac{r'^3\,dr'}{a^2-r'^2}=\frac{8a^3}{3}\mu_0 H=\frac{d^3}{3}(\mu_0\epsilon_0)^{1/2}E$$

where $d\ (=2a)$ is the diameter of the disc.

The magnitude of the scattered field back in the incident directions due to a magnetic dipole of moment M is

$$E_s=\frac{\pi}{(\mu_0\epsilon_0)^{1/2}\lambda^2 r}M=\frac{\pi d^3}{3\lambda^2 r}E.$$

The fields due to the electric and magnetic dipoles are in the same direction and phase and therefore additive. Hence the total scattered field when the disc is in the plane of incidence and normal to the magnetic vector is $E_s=\pi d^3E/\lambda^2 r$ and the equivalent echoing area is

$$\sigma=\frac{4\pi r^2|E_s|^2}{|E|^2}=\frac{4\pi^3 d^6}{\lambda^4}.$$

(*c*) The validity of the approximations involved are discussed in the detailed treatments by Bethe (1944) and Copson (1946). A more rigorous analysis has been made by Bouwkamp (1954) by solving directly, in general form, the electromagnetic equations satisfying the boundary conditions. He finds that the solution contains not only the results derived in (*a*) and (*b*) above but also introduces further closed-loop currents. These consist of two symmetrically disposed current systems circulating in opposite directions. The magnitude of these currents is less than that of the circular currents determined in (*b*) above by a factor of the order of a/λ. For this reason, and also because they form two equal and opposite adjacent magnetic dipoles, their contribution to the far field is negligible. The equivalent echoing area is thus unaffected by these additional currents required for the more precise solution of the boundary conditions.

REFERENCES

BETHE, H. A., 1944, *Phys. Rev.*, **66**, 163.

BOUWKAMP, C. J., 1954, *Rep. Progr. Phys.*, **17**, 35 (London : Physical Society).

COPSON, E. T., 1946, *Proc. Roy. Soc.* A, **186**, 100.

KERR, D. E., 1951, *Propagation of Short Radio Waves*, M.I.T. Radiation Series, 13 (New York : McGraw-Hill).

MIE, G., 1908, *Ann. Phys., Lpz.*, **25**, 377.

STRATTON, J. A., 1941, *Electromagnetic Theory* (New York : McGraw-Hill).

RESEARCH NOTES

Saturation Magnetization in Copper–Nickel Alloys.

By S. A. AHERN† and W. SUCKSMITH
Department of Physics, University of Sheffield

MS. received 30*th July* 1956

The variation of the saturation magnetization $\sigma_{0,0}$ at the absolute zero of temperature with composition in ferromagnetic alloy systems is an important aid to a knowledge of the energy bands in the elements of the first transition series.

The system presenting the greatest simplicity is that of copper and nickel. Investigations of the magnetic properties in the ferromagnetic range of this system have been carried out previously by Alder (1916), but as it was felt that the accuracy of measurement required for adequate interpretation could be improved upon, it was decided recently to institute a detailed programme of measurements into the ferromagnetic properties of this system, the programme to include measurements of the magnetization–temperature curves for each alloy over as great a range of temperature as possible, together with measurements of the magnetocaloric effect in the Curie point region, this being the most reliable method of obtaining the Curie temperature. A description of the construction of the apparatus and the initial measurements has been given by Oliver and Sucksmith (1953). It is the result of subsequent measurements made to obtain values of the saturation magnetization for a total of six alloys with atomic compositions ranging from 4·7% of copper to 34·3% of copper on which it is desired to report here.

Although the results are incomplete owing to a delay in obtaining measurements at liquid hydrogen temperatures, it is thought that the data which have already been obtained are of sufficient importance to record.

Values of Saturation Moment and Curie Point, etc.

Atomic Composition (% Cu)	Ferromagnetic Curie Point (°K)	Saturation Moment (g^{-1})	Bohr Magnetons
34·3	291	21·4	0·231
30·8	310 303‡	23·1	0·249
18·9	420	35·7	0·381
16·0	—	40·5	0·431
8·6	541	48·3	0·511
4·7	587	52·5	0·554
(0)	(631)	(57·6)	(0·606)

‡ Obtained using measurements of magnetocaloric temperature changes.

The values of the parameters which have been measured are shown in the table. The values of $\sigma_{0,0}$ were obtained by means of a T^2 extrapolation of the spontaneous magnetization to the absolute zero of temperature. Values of the spontaneous magnetization in this region, well below the Curie point, were obtained by extrapolation of the magnetic isothermals, plotted for fields of between 5000 and

---

† Now at the Pennsylvania State University, University Park, Pennsylvania, U.S.A.

20 000 Oe. The Curie point values were obtained by an extrapolation of the square of the spontaneous magnetization to the temperature axis, the intercept being taken as the Curie temperature. In the case of one of the alloys noted in the table, an additional value for the Curie temperature has been obtained from magnetocaloric measurements. In figures 1 and 2 respectively the saturation moment and Curie temperature are plotted as a function of composition for comparison with Alder's results (indicated by the dotted lines). One point taken from the results of Oliver and Sucksmith is included at 23·9% copper. It will be seen from figure 1 that the straight line drawn through nickel and the first five points intersects the composition axis at approximately 52%. If the final point is considered in conjunction with the results of Arrott (1954) on alloys containing 58 and 63% of copper, it further appears that the linear relation of both saturation moment and Curie point with composition breaks down at about 30% of copper.

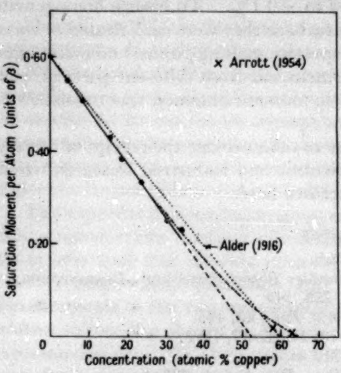


Figure 1. Variation of moment per atom (in Bohr magnetons) with concentration.

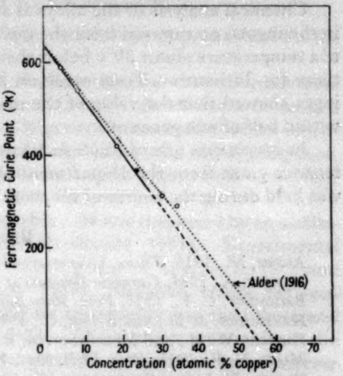


Figure 2. Variation of ferromagnetic Curie point with concentration.

Alder's results have been interpreted by Mott (1935) on the basis that each copper atom alloyed with nickel contributes one electron to a common 3d–4s band system, raising the common Fermi level in such a way that the slope of the resultant magnetic moment–composition curve is unity. The initial slope given by the present results indicates that on this basis each copper atom contributes more than one electron, which is unlikely. In this connection, however, if we take $g = 2{\cdot}2$, as obtained by ferromagnetic resonance experiments (Bagguley 1955), we have $\mu = g\frac{1}{2}\beta = 1{\cdot}1\beta$. The breakaway from the straight line relation for the 34·3% alloy taken in conjunction with Arrott's results mentioned above is much less disturbing. In fact, this is qualitatively consistent with the low temperature specific heat measurements of Keesom and Kurrelmeyer (1940) and measurement of temperature variation of susceptibility for alloys containing more than 60% of copper (Wheeler 1939, Kaufmann and Start 1943).

As the measurements discussed are incomplete, it is more important to discuss the accuracy to be expected from the measurements which have been made, particularly in the case of those alloys with the higher copper concentrations, where the extrapolation to the absolute zero of temperature is rather extended.

The spontaneous magnetization values are obtained from the extrapolated magnetic isothermals with an estimated accuracy of ±0·5%. Assuming that the T^2 extrapolation of moment against temperature is valid down to the absolute zero, then the maximum possible error in $\sigma_{0,0}$ is ±2·0%, yet to make the present data lie on the line which passes through the nickel point and extrapolates to zero at 60% of copper, the values of the saturation moment must be increased by up to 10%.

Chemical analysis of the alloys is stated to ±0·1%. To ensure homogeneity in the ingots, on removal from the melting furnace they were each heated *in vacuo* to a temperature about 30°c below their respective melting points, and maintained there for 36 hours. Tests made on specimens cut from different parts of each ingot showed that the value of the magnetic moment obtained was consistent to within half of one per cent.

In conclusion one of us (S.A.A.) desires to acknowledge the receipt of a maintenance grant from the Department of Scientific and Industrial Research which was held during the course of the work described here.

References

Alder, M., 1916, *Thesis*, University of Zürich.
Arrott, A., 1954, *Carnegie Institute of Technology Report*. Pittsburgh, Pennsylvania.
Bagguley, C. F., 1955, *Proc. Roy. Soc.* A, **228**, 549.
Kaufmann, A. D., and Starr, C., 1943, *Phys. Rev.*, **63**, 445.
Keesom, W. H., and Kurrelmeyer, B., 1940, *Physics*, **7**, 1003.
Mott, N. F., 1935, *Proc. Phys. Soc.*, **47**, 571.
Oliver, D. J., and Sucksmith, W., 1953, *Proc. Roy. Soc.* A, **291**, 1.
Wheeler, M. A., 1939, *Phys. Rev.*, **56**, 1137.

Spectroscopic Splitting Factors for Iron and Silicon Iron

By G. S. BARLOW and K. J. STANDLEY
Department of Physics, University of Nottingham

Communicated by L. F. Bates; MS. received 27th April 1956

Microwave resonance absorption has been studied in polycrystalline iron and 2·5% silicon–iron specimens, as well as in a single crystal of the latter material, at room temperature and at wavelengths close to 1·227 and 0·864 cm. Disc-shaped samples about 11 mm diameter and 0·25 mm thick were used. The longer wavelength apparatus and method of measurement have already been described (Adam and Standley 1953, Standley and Reich 1955); the higher frequency equipment differed only in detail and does not warrant further description here.

In experiments of this nature, made at low power levels and at a fixed frequency ν, the losses in the specimen rise to a maximum in an applied magnetic field $H_z^{\max}$. For a single crystal, Kittel (1948) and others have shown that

$$h\nu = g\beta\{[H_z^{\max} + (N_y - N_z)M + \phi(K_1, K_2, M, \theta)][H_z^{\max} + (N_x - N_z)M + \psi(K_1, K_2, M, \theta)]\}^{1/2}. \qquad \ldots\ldots(1)$$

Here g is the spectroscopic splitting factor, β is the Bohr magneton, N_x, N_y, N_z are the demagnetization factors in the three coordinate directions, and M is the magnetization in the specimen which is assumed to be saturated in the field $H_z^{\max}$. ϕ and ψ are algebraic functions which take account of the magnetocrystalline anisotropy of the specimen. For a cubic crystal, when the surface of the specimen is a (110) face, as in our measurements, Healy (1951) has shown that

$$\begin{aligned}\phi &= (1-2\sin^2\theta-\tfrac{3}{8}\sin^2 2\theta)\,2K_1/M-\sin^2\theta\cos^2\theta\,(3\sin^2\theta+2)\,K_2/2M\\ \psi &= (2-\sin^2\theta-3\sin^2 2\theta)\,K_1/M\\ &\quad +\sin^2\theta\,(6\cos^4\theta-11\sin^2\theta\cos^2\theta+\sin^4\theta)\,K_2/2M.\end{aligned}$$

Here K_1 and K_2 are the first and second order anisotropy constants and θ is the angle between H_z and the cube edge.

When a polycrystalline specimen is used, random orientations of the individual crystallites are assumed, and an averaging process must be carried out to determine the conditions for maximum absorption. It is generally assumed that the resonance equation is that given by (1) above, the average values of ϕ and ψ being zero. This appears to be an approximation which may lead to appreciable errors in g under certain conditions and it is briefly discussed below.

The experimental results are given in the table; M was measured by an oscillation magnetometer method (Griffiths and Macdonald 1951). Experimental errors were such that absolute values of $\mu_R^{\max}$ and g may be in error by 15% and 1·5–2% respectively (Standley and Reich 1955). Relative errors between measurements at the two wavelengths should be appreciably less than these and *relative* values of g should be reliable to $\pm 0{\cdot}02$. Three points emerge from the experimental data: (i) The g-values for polycrystalline specimens tend to increase with decreasing wavelength. This trend confirms the results of Bagguley (1955) who reported 2·00 and 2·06 at wavelengths of 3·14 and 1·25 cm respectively, for both iron and silicon iron polycrystalline specimens. (ii) The g-values for the single crystal appear to diminish with decreasing wavelength. (iii) The maximum values of the apparent permeability $\mu_R^{\max}$ decrease with decrease in wavelength. This is also true of Bloembergen's (1950) measurements on nickel and supermalloy at 3·0 and 1·25 cm.

Measured g-values which are frequency (or field) dependent are usually regarded as deriving from incorrect interpretations of experimental data, as it is expected that the true spectroscopic splitting factor will be a constant for a given material. Possible explanations of (i) and (ii) were therefore sought.

Consider first the polycrystalline results; Miles (1954) semi-qualitatively and Bagguley (1955) quantitatively for a special case suggested that the correct resonance equation for polycrystalline materials is not obtained by equating ϕ and ψ to zero in equation (1). Standley and Stevens (1956) extended Bagguley's calculation and showed for spherical specimens that there is a shift of the resonance peak of magnitude dependent upon the width of the single crystal line. The shift approaches $K_1/2M$ for narrow lines and differs little from this value for lines such as those found in iron single crystals. The shift is to higher values of H_z for positive K_1, as in iron, i.e. too small an apparent g-value is calculated if no correction is applied. Although the more involved computation for a disc-shaped specimen has not been carried out, it is reasonable to expect a similar order of magnitude and certainly the same sign of the correction. For iron $K_1/2M \sim 100$ oersteds and at 1·23 cm $H_z^{\max} \sim 3000$ oersteds. At this wavelength, the uncorrected g-value might be about 2% too small, while at 0·87 cm the correction would probably be within the relative error of $\pm 0{\cdot}02$.

This correction will not apply, of course, to the single crystal measurements. With these, and with the polycrystalline measurements also, it is necessary to consider the effect upon the resonance condition of the finite penetration of the microwave field into a planar specimen. Kittel and Herring (1950) first attempted the calculation but this was recently criticized by Rado and Weertman (1954) and Ament and Rado (1955). The resonance equation derived by the last named authors is applicable, unfortunately, only to longer wavelength measurements on easily saturated materials when, for example, $H_z^{\max}$ can be neglected with respect to $4\pi M$. The form of their equation is such that a shift of the resonance peak to lower values of H_z is predicted and if, as Ament and Rado note, the Kittel and Herring formula is used as a rough approximation, in calculating g-values we must add to $H_z^{\max}$ a term $a\mu_R^{\max}$ where a is a positive constant for the material. Since $\mu_R^{\max}$ decreases with decreasing wavelength this form of correction term could give a constant g value when applied to the experimental single crystal data. Moreover, the Kittel and Herring formula predicts a shift of about 30 oersteds in silicon-iron at 1·23 cm which is of the correct order of magnitude.

The theoretical equation (1) should be modified to take account of shifts of this kind. Fortunately, we found the anisotropy constants obtained by fitting experimental single crystal data to this equation were not critically dependent upon absolute peak location.

We thus found for the silicon–iron single crystal at 1·23 cm

$$K_1 = +(3{\cdot}8 \pm 0{\cdot}1)\times 10^5,\; K_2 = +(1{\cdot}3 \pm 0{\cdot}2)\times 10^5 \text{ ergs cm}^{-3}$$

while at 0·87 cm the values were

$$K_1 = +(3{\cdot}6 \pm 0{\cdot}1)\times 10^5,\; K_2 = +(1{\cdot}4 \pm 0{\cdot}2)\times 10^5 \text{ ergs cm}^{-3}.$$

From the foregoing discussion, it is apparent that more reliable g-values are obtained by the extrapolation of measurements extending to the highest available frequencies, preferably using single crystal specimens. The experimental single crystal data combine to suggest a probable g-value for iron of $2{\cdot}09 \pm 0{\cdot}02$. Within experimental error iron then obeys the predicted equation (Kittel 1949, Van Vleck 1950) $g-2=2-g'$ since direct gyromagnetic measurements (see for example Brown *et al.* 1954) lead to a value of $g' = 1{\cdot}93 \pm 0{\cdot}01$.

Material	M (e.m.u. cm^{-3})	λ (cm)	$\mu_R^{\max}$	Apparent g-value
Polycrys. Iron	1720	1·227	70	2·04
	1720	0·864	—	2·10
Polycrys. 2·5% Si–Fe	1630	1·227	100	2·03
	1630	0·864	60	2·06
Single Crystal 2·5% Si–Fe	1630	1·226	120	$2{\cdot}13_8$
	1630	0·869	94	$2{\cdot}10_5$

ACKNOWLEDGMENTS

The authors wish to express their thanks to Professor L. F. Bates for his continued interest and advice and to Dr. K. W. H. Stevens for several valuable discussions. One of us (K. J. S.) is indebted to the Government Grant Committee of the Royal Society for a grant which made possible the purchase of the 0·8 cm microwave apparatus, while the other (G. S. B.) wishes to thank the Department of Scientific and Industrial Research for a maintenance award.

REFERENCES

ADAM, G. D., and STANDLEY, K. J., 1953, *Proc. Phys. Soc.* A, **66,** 823.
AMENT, W. S., and RADO, G. T., 1955, *Phys. Rev.*, **97,** 1558.
BAGGULEY, D. M. S., 1955, *Proc. Roy. Soc.* A, **228,** 549.
BLOEMBERGEN, N., 1950, *Phys. Rev.*, **78,** 572.
BROWN, S., MEYER, A. J. P., and SCOTT, G. G., 1954, *C.R. Acad. Sci., Paris,* **238,** 2504.
GRIFFITHS, J. H. E., and MACDONALD, J. R., 1951, *J. Sci. Instrum.*, **28,** 56.
HEALY, D. W., 1951, *Harvard University, Cruft Laboratory,* Tech. Rep. 135.
KITTEL, C., 1948, *Phys. Rev.*, **73,** 155; 1949, *Ibid.*, **76,** 743.
KITTEL, C., and HERRING, C., 1950, *Phys. Rev.*, **77,** 725.
MILES, P. A., 1954, *Nature, Lond.*, **174,** 177.
RADO, G. T., and WEERTMAN, J. R., 1954, *Phys. Rev.*, **94,** 1386.
STANDLEY, K. J., and REICH, K. H., 1955, *Proc. Phys. Soc.* B, **68,** 713.
STANDLEY, K. J., and STEVENS, K. W. H., 1956, *Proc. Phys. Soc.* B, **69,** 993.
VAN VLECK, J. H., 1950, *Phys. Rev.*, **78,** 266.

The Origin of the Hydrated Oxide formed when Anodic Aluminium Oxide Films are ' Sealed ' in Boiling Water

BY I. S. KERR †

Applied Physical Chemistry Laboratory, Chemical Engineering Department,
Imperial College, London

MS. received 18*th May* 1956

IT is widely known that the freshly anodized aluminium surface gives a halo electron-diffraction pattern, whereas after sealing by immersion in boiling water a pattern corresponding to boehmite, γ-AlOOH, is obtained (Phillips 1952). In recent years several authors have investigated the origin of the boehmite and most have concluded that it must occur through further hydrated oxidation of aluminium.

Thus, Harrington and Nelson (1940) obtained electron-diffraction patterns due to boehmite from sealed specimens. After abrading on 4/0 emery the original halo pattern was restored. By making use of the limited depth of penetration of electrons, Hart (1954) was able to estimate the rate of growth of boehmite on anodized electro-polished single-crystal surfaces of aluminium. This rate of corrosion proved to be linear, from which it was concluded that the amorphous layer took no part in the reaction other than transport of ions.

This note describes further new evidence on the origin of the γ-AlOOH, made possible by the ability of electron microscopes to obtain diffraction patterns from particular areas of the specimen (Matthews and Wilman 1955). The anodized films used for this purpose were prepared by Dr. A. Charlesby at the Atomic Energy Research Establishment, Harwell, and consisted of aluminium foil (about 0·01 cm thick) anodized at 50 v in an ammonium borate solution. Most of the aluminium was then stripped from the oxide with chlorine at 200°C, leaving a metal foil network sufficiently rigid to support the unaffected oxide film across 'windows' from which the metal was completely removed. As the oxide film thickness was about 700 Å (14 Å per volt) it was possible to penetrate it with an electron beam without unduly large incoherent scattering.

† Now at the Chemistry Department, Imperial College, London

Before sealing the film was examined in the electron microscope. When suitable 'windows' were found and the electron microscope was adjusted for diffraction, the expected halo pattern was obtained.

The foil was then sealed by immersion in boiling distilled water for 20 minutes and returned to the microscope. Certain areas gave a halo pattern (figure 1 (*a*), Plate) identical with the original, while other areas gave the sharper boehmite pattern (figure 2 (*a*), Plate).

Closer examination proved that the halo patterns originated from the centre of large areas of oxide film, or from tongues which projected from the metal but had otherwise free edges, as seen in figure 1 (*b*). The boehmite pattern however, came from parts of the oxide film near the edges of the metal or from regions where small specks of aluminium were dispersed over the film, as shown in figure 2 (*b*). The texture of the two areas of film also appears markedly different.

It is thus concluded that the boehmite is formed only in the immediate neighbourhood of aluminium.

ACKNOWLEDGMENTS

The author thanks Mr. H. I. Matthews and Dr. H. Wilman of this laboratory, and Dr. A. Charlesby of the Metallurgy Division, Atomic Energy Research Establishment, Harwell, for their help.

REFERENCES

HARRINGTON, R. A., and NELSON, H. R., 1940, *Amer. Inst. Min.* (*Metall.*) *Engrs;* Inst. Metals Div. **137**, 62, *Tech. Pub.* 1158.
HART, R. K., 1954, *Trans. Faraday Soc.*, **50**, 269.
MATTHEWS, H. I., and WILMAN, H., 1955, *Brit. J. Appl. Phys.*, **6**, 277.
PHILLIPS, H. W. L., 1952, *Monogr. Scr. Inst. Metals, No.* 13, p. 237.

Defects with Several Trapping Levels in Semiconductors

BY P. T. LANDSBERG
Marischal College, University of Aberdeen

MS. received 10*th May* 1956

DEFECTS in semiconductors which exhibit several trapping levels are by now well known: examples are cobalt, copper and gold in germanium, and thermally induced defects in silicon (cf. Burton 1954 for references). In such cases the usual methods of determining the position of the Fermi level are inadequate. A correct general discussion which may be adapted to these situations is not available in the literature, but has been referred to by Brooks (1955). It may therefore be worth while to give the details of this treatment. The difficulty which arises is due to the following obvious fact. Since the defects can be singly or doubly ionized, they can exhibit two different spectra of trapping levels; the number of levels of each type is therefore not a constant, but depends on the average degree of single and double ionization of defects in the crystal (i.e. on the temperature). The trapping levels which occur are therefore no longer independent, as in simple semiconductor theory.

The general problem is as follows. Consider N identical impurity centres, distinguishable in virtue of their position, and suppose that each centre can trap at most M electrons. The supposed inability of a centre to trap an $(M+1)$th electron is part of our model of the centre. If desired, one may regard this restriction as an idealization of a situation in which the system which consists of the centre plus r electrons $(r>M)$ has such a high energy that situations of this type may be neglected. If a centre has trapped j electrons, it will be assumed that these can be in any one of the j-electron quantum states of the centre. The energy of the ith j-electron state will be denoted by $E(i,j)$. It will be assumed that the trapping processes at different centres are independent of each other. The grand partition function of the system of N centres is

$$\text{g.p.f.}=\left[\sum_{j=1}^{M}\lambda^j Z_j\right]^N, \qquad \ldots\ldots(1)$$

where

$$Z_j\equiv\sum_i \exp\left[-\frac{E(i,j)}{\boldsymbol{k}T}\right],\quad \lambda\equiv\exp\left(\frac{\mu}{\boldsymbol{k}T}\right).$$

λ is the absolute activity, μ the chemical potential or Fermi level, Z_j the ordinary partition function of a centre which has trapped j electrons, and the i-summation extends over all j-electron states of a centre. The mean number of electrons trapped on the N centres is $n=\lambda\{\partial[\log \text{g.p.f.}]/\partial\lambda\}_{V,T}$, so that

$$\frac{n}{N}=\left[\sum_0^M j\lambda^j Z_j\right]\Big/\left[\sum_0^M \lambda^j Z_j\right]. \qquad \ldots\ldots(2)$$

This may be written in a form proposed earlier (Landsberg 1952),

$$\frac{n}{N}=\frac{1}{1+\beta\exp[(\epsilon-\mu)/\boldsymbol{k}T]}, \qquad \ldots\ldots(3)$$

provided β is defined by

$$\beta\equiv\left[\exp\left(-\frac{\epsilon}{\boldsymbol{k}T}\right)\right]\left[Z_0-\sum_1^M(j-1)\lambda^j Z_j\right]\Big/\left[\sum_1^M j\lambda^{j-1}Z_j\right]. \qquad \ldots\ldots(4)$$

ϵ is an arbitrary energy, which cancels out of (3). Finally, (3) may be reduced, at least formally, to the usual Fermi function, by associating with each centre a single effective trapping level of energy $E'=\epsilon+\boldsymbol{k}T\log\beta$. The simplicity of the resulting form

$$n/N=\{1+\exp[(E'-\mu)/\boldsymbol{k}T]\}^{-1} \qquad \ldots\ldots(5)$$

of (3) is, however, only apparent. For, if $M\geqslant 2$, E' involves the Fermi level, and is therefore dependent on temperature and impurity concentration. For this reason use of a single energy level for a centre which has in fact a set of trapping levels, leads to a spurious temperature and concentration dependence of this level.

The equation for the Fermi level, assuming $Z_0, Z_1, \ldots, Z_M$ to be known, is

$$\frac{4\pi}{\boldsymbol{h}^3}(2m^*\boldsymbol{k}T)^{3/2}\int_0^\infty\frac{x^{1/2}\,dx}{1+\lambda^{-1}\exp x}=N_e-\left[N\sum_0^M j\lambda^j Z_j\right]\Big/\left[\sum_0^M\lambda^j Z_j\right] \qquad \ldots\ldots(6)$$

in a usual notation. N_e is the number of electrons available in donors and conduction band at the absolute zero. (6) assumes that holes need not be considered, otherwise (6) would have to be generalized by the introduction of a second Fermi integral. It is also assumed that only one type of centre is present, since each type of centre contributes a fraction such as that shown on the right-hand side of (6).

For a non-degenerate semiconductor (6) represents an equation of order $M+1$ in λ.

Examples

(A) If a centre can trap only one electron,

$$\frac{n}{N}=\frac{\lambda Z_1}{Z_0+\lambda Z_1}. \qquad \ldots\ldots(7)$$

(B) If, in addition, the centre without the electron has an a-fold degenerate state of energy E_0, and the centre with the electron has a b-fold degenerate state of energy E_0+E, then

$$\frac{Z_0}{Z_1}=\frac{a\exp[-E_0/\boldsymbol{k}T]}{b\exp[-(E_0+E)/\boldsymbol{k}T]}=\frac{a}{b}\exp\left(-\frac{E}{\boldsymbol{k}T}\right), \text{ i.e. } \beta=\frac{a}{b}. \qquad \ldots\ldots(8)$$

If $a=1$, $b=2$, or $a=2$, $b=1$, then the cases of ***unpaired*** and ***paired*** electrons are obtained respectively.

(C) If a centre can trap two electrons (e.g. gold in germanium, Dunlap 1955), two distinct energy levels E_1, E_2 are available to the first electron, and two energy levels E_1', E_2' to the second electron. Here E_1' is the energy to which E_1 has been raised as a result of E_2 being occupied, and E_2' is the energy to which E_2 has been raised as a result of E_1 being occupied. The energy of the state in which two electrons are trapped is, however, independent of the order in which the states were occupied, and will be denoted by E_3. If the g's be appropriate degeneracies, we have

$$Z_1=\sum_{j=1}^{2} g_j\exp(-E_j/\boldsymbol{k}T), \qquad Z_2=g_3\exp(-E_3/\boldsymbol{k}T). \qquad \ldots\ldots(9)$$

If we neglect the effect of the higher level in Z_1, and take N_e as $2N$, the right-hand side of (6) becomes

$$\frac{[2Z_0+g_1\lambda^1\exp(E_1/\boldsymbol{k}T)]\,N}{Z_0+g_1\lambda^1\exp(-E_1/\boldsymbol{k}T)+g_3\lambda^2\exp(-E_3/\boldsymbol{k}T)}, \qquad \ldots\ldots(10)$$

and hence the Fermi level can be obtained, if reasonable estimates are available for Z_0, g_1, g_3, E_1 and E_3 (Dunlap's results suggest $E_1=-0{\cdot}60$ ev, $E_3=-0{\cdot}20$ ev).

(D) In this last example an analogy between semiconductivity and adsorption will be pointed out. The above theory remains valid if impurity centres are interpreted as adsorption sites, and trapped electrons as adsorbed atoms or molecules. As an example, note that if the gas is regarded as ideal, then $\lambda=p\boldsymbol{h}^3/(2\pi\boldsymbol{m}\boldsymbol{k}T)^{3/2}\boldsymbol{k}Tj(T)$, where p is the pressure of the gas phase, $j(T)$ the internal partition function of the particles of the gas, and the other symbols have their usual meanings. Since the partition function Z_1 is independent of pressure, one can define a pressure-independent quantity p_0 by $\lambda Z_1=p/p_0$. If no more than one particle can be adsorbed by a site, (7) applies, and yields now

$$c=p/(p_0+p), \qquad \ldots\ldots(11)$$

where n/N has been interpreted as the *fractional coverage* of the surface. This is the well-known Langmuir adsorption isotherm for the adsorption of particles at localized sites, the number of particles per site being zero or one. Equations (7) and (11) are in exact correspondence. Other adsorption isotherms can also be brought into correspondence with trapping mechanisms at impurity centres, but (7) and (11) seem at present to represent the most important example of a situation which corresponds to a physically meaningful model in both semiconductivity and adsorption.

REFERENCES

BROOKS, H., 1955, *Advances in Electronics and Electron Physics* (New York : Academic Press), p. 127.
BURTON, J. A., 1954, *Physica*, **20**, 845.
DUNLAP, W. C., 1955, *Phys. Rev.*, **97**, 614.
LANDSBERG, P. T., 1952, *Proc. Phys. Soc.* A, **65**, 604.

Stacking Faults in Vacuum Evaporated Crystals of Copper and Silver

BY T. B. RYMER, F. J. FAYERS AND S. J. HEWITT
Physics Department, University of Reading

MS. received 11*th May* 1956

PRECISION electron diffraction measurements of the interplanar spacings of crystals in vacuum evaporated copper films have been made by Halliday, Rymer and Wright (1954). Deviations from the normal spacings were attributed to uniform stresses throughout the specimen. The breadth of the diffraction rings was attributed to the finite size of the crystallites, which were assumed to be discs approximately 90 Å diameter and 15 Å thick parallel to 111 planes. Garrod (1956) has pointed out that closer agreement with Halliday, Rymer and Wright's measurements is obtained if it be assumed that in addition to the system of stresses postulated by these authors there are also stacking faults on the 111 planes. Garrod also suggested that the existence of these stacking faults, and not the finite size of the crystallites, is responsible for the breadths of the diffraction rings.

Garrod has given the following expression for the fractional deviation δ of the spacing of the planes *hkl* from the normal value :

$$-\delta = \tfrac{1}{2} PM \qquad \ldots\ldots(1)$$

where

$$M = a\,|h+k+l|/(\Sigma h^2)$$

and $a = +1, 0, -1$ according as $|h+k+l| = 3N+1, 3N, 3N-1$, N being an integer. P is related to the probability of faulting ϵ :

$$P = \frac{\sqrt{3}}{2\pi}\,\epsilon \text{ for deformation faults,} \qquad \ldots\ldots(2a)$$

$$P = \frac{\sqrt{3}}{2\pi}\left(\frac{\epsilon}{1-\epsilon}\right)^2 \text{ for growth faults.} \qquad \ldots\ldots(2b)$$

It is assumed that the radius of a diffraction ring formed by the superposition of equivalent *hkl* reflections corresponds to the mean value of δ for all equivalent *hkl*. This, of course, means that the peak of a composite diffraction ring is the centroid of the peaks of the components when these are given equal weight. This assumption is correct provided the components have the same width. For components of unequal width, the resultant peak is the centroid of the peaks of the components when each is weighted as the inverse cube of its width (Rymer 1956). Garrod's application of the stacking fault theory is therefore valid only if all equivalent *hkl* reflections have the same width. This is not the case if the widths are due to the stacking faults themselves, and therefore this interpretation of the widths must be rejected.

The assumption of Halliday, Rymer and Wright that the widths of the rings are due to the finite size of the crystallites also leads to unequal widths for equivalent reflections and, if true, would invalidate the stacking fault theory. In fact, later work has cast doubt on this interpretation of the widths. Using a high resolution electron diffraction camera having a focused spot size of 5–10 μ, together with a microphotometer with a 5 μ slit width, a study has been made of the contours of diffraction rings of copper. It has been found that these are accurately gaussian (figure 1). A detailed theoretical study of the factors influencing line contours (to be reported in a future paper) shows that such a contour cannot be produced by particle size broadening, but is readily explained by assuming a random distribution of strains such as would be produced by forces exerted by crystallites on one another when packed together. From the slope of the line of figure 1 the root mean square value of the strains perpendicular to the reflecting plane can be calculated. This is given in the table, together with values determined by similar measurements on other diffraction rings from the same specimen. From the elastic constants of copper (Boas and Mackenzie 1950), Young's modulus in directions perpendicular to different planes has been calculated and recorded in the third row of the table. There is a tendency for the strains to be larger in directions where Young's modulus is smaller. This is to be expected if the strains are due to random intercrystalline forces, and supports the view that the breadth of the different rings is due to random stresses, and therefore that equivalent *hkl* reflections will have the same width.

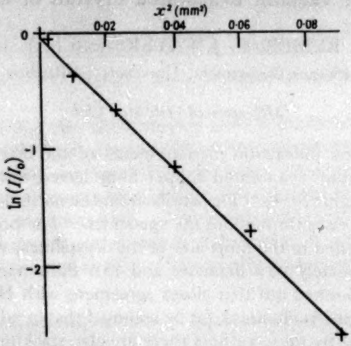


Figure 1. Contour of 200 reflection from copper. The logarithm of the intensity of the line is plotted against the square of the distance x from the peak of the line.

Young's Modulus and Root Mean Square Strains in direction *hkl*

hkl	111	200	220	311	331	420
strain × 10^3	8·6	12·0	10·8	9·4	6·4	6·6
Young's modulus × 10^{-11} (dyn cm^{-2})	19·1	6·7	13·0	9·6	14·4	9·7

Garrod's theory of stacking faults, together with his implicit assumption that equivalent reflections have the same width, is inherently reasonable and may well be true for films of vacuum evaporated copper. Such films do not, however, provide a very good demonstration of stacking faults. Most of the displacement of the diffraction rings is due to uniform stresses acting throughout the specimen as a whole, in accordance with the theory of Halliday, Rymer and Wright. Only a small part of the total displacement can be attributed to stacking faults. It would obviously be of interest to obtain diffraction patterns corresponding to stacking faults in the absence of uniform stress. We have obtained such patterns from vacuum evaporated silver specimens. Figure 2 shows the results of fitting the stacking fault theory of equations (1) and (2) to measurements of such a pattern evaluated as in the papers referred to. The agreement is within the limits of experimental error. We therefore conclude (*a*) that the crystals in evaporated silver specimens contain stacking faults, (*b*) that the breadth of the rings is due to random strains produced by the forces which the crystals exert on one another, (*c*) that there is no detectable stress acting throughout the specimen as a whole. From the slope of the line in figure 2, $P = 0{\cdot}015$. If the faults are due to deformation (equation (2*a*)), the probability of faulting $\epsilon = 6\%$; for growth faults (equation (2*b*)), $\epsilon = 19\%$. Since the breadth of the rings ($\sim 500\,\mu$) implies strains of the order of $1\frac{1}{2}\%$ and therefore large stresses of the order of 10^{10} dyn cm^{-2} between the crystals, it seems reasonable to suppose we are dealing with deformation rather than growth faults.

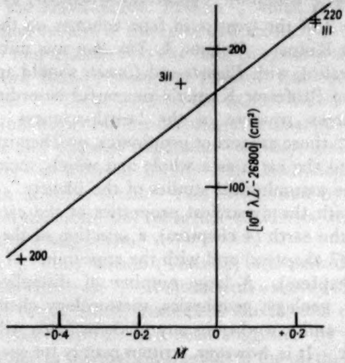


Figure 2. Evaporated silver. Comparison of results with stacking fault theory. λ and L are the apparent electron wavelength and specimen plate distance. The ordinate of the figure is (const.-26800δ) where the constant is of the order of 100.

References

Boas, W., and Mackenzie, J. K., 1950, *Progr. Metal Phys.*, **2**, 90.
Garrod, R. I., 1956, *Proc. Phys. Soc.* B, **69**, 260.
Halliday, J. S., Rymer, T. B., and Wright, K. H. R., 1954, *Proc. Roy. Soc.* A, **225**, 548.
Rymer, T. B., 1956, *Proc. Roy. Soc.* A, **235**, 274.

REVIEWS OF BOOKS

The Earth as a Planet, edited by GERARD P. KUIPER. Pp. xvii + 751. (Chicago: University Press; London: Cambridge University Press, 1954.) £4 14*s*.

This is Volume 2 of the treatise in four volumes on the *The Solar System* edited by Professor Kuiper. Volume 1, *The Sun* was published in 1953, and volumes 3 and 4 dealing with *Planets* and *Comets* should appear shortly. The book is the result of Professor Kuiper's successful co-ordination of the efforts of sixteen well-known workers in the 'earth-sciences' from six different countries, to cover "those aspects of geophysics, geochemistry and atmospheric physics as pertain to the earth as a whole and which, incidentally, are able to assist and guide the astrophysical studies of the planets". It contains fifteen chapters, dealing with the dynamical properties of the earth (2 chapters), the bulk properties of the earth (4 chapters), a selection of the manifold problems of the atmosphere (7 chapters) and with the appearance of the earth as 'seen' from outside (2 chapters). A large number of disciplines are embraced—physics, astronomy, geology, geophysics, meteorology chemistry, biochemistry and oceanography—and a worker in any of these fields will find something of interest in this book. It is, however, written mainly for specialists and contains numerous references. Although many of the articles were first written in 1952, the authors were able to revise their contribution just before going to press, thereby bringing them up to date. This is an important feature of a book which tends to give pride of place to contemporary work in a field where progress is rapid.

At the present day we have a fair knowledge of the planet on which we live, and there is almost complete agreement on what is believed to be its general structure. The earth is a nearly spherical body rather more than 600 km in radius. Although the circumstances of its birth are obscure, this event probably occurred about 4×10^9 years ago. It rotates at an *approximately* uniform rate, once every twenty-four hours, and about it revolves its one satellite, the moon.

Below a depth from the surface of about 30 km the earth is approximately spherically symmetrical in its structure. It contains a large liquid metallic core, the temperature of which is probably several thousands of degrees centigrade. The core is surrounded by a deep solid mantle, composed mainly of the mineral olivine, a ferromagnesium silicate. Overlying the mantle is a thin crust, with the study of which geology is mainly concerned, and surrounding the earth is its atmosphere.

Forty years ago when "geological maps stopped at the coasts and told little of three quarters of the earth's surface and practically nothing of the interior" (p. 139), and when only comparatively rudimentary studies of the atmosphere were made, a large part of what was known about the earth might well have been treated in a book of seven hundred pages. However, progress in the earth-sciences has been very rapid during the past four decades and as a result, Professor Kuiper's problem of selecting suitable topics must have been a very difficult one. Although his choice was influenced to some extent by the recent publication of several comprehensive treatises each dealing in detail with some special topic, the atmosphere is given a prominent place in the book because " the study of planetary atmospheres is at present one of the most hopeful approaches towards a fuller understanding of the planets and their evolutions ".

Chapter 1, entitled " Dimensions and Rotation " is by the Astronomer Royal, Sir Harold Spencer Jones, who deals with the figure, mass and rotation of the earth. This is followed by Sir Harold Jeffreys' (Cambridge) article on the " Dynamics of the Earth–Moon System ", which is a short essay on the changes of the rotation vectors of the earth and moon.

The first of the four chapters on the bulk properties of the earth is a long article by Sir Edward Bullard (National Physical Laboratory) entitled " The Interior of the Earth ". Those parts of the earth " below the levels at which the geologist, the mineralogist and the petrologist ordinarily work " are discussed. First the properties and physical conditions—density, composition, ellipticity, pressure and gravity—as inferred mainly from seismological data and earth tides, are dealt with in some detail. This is followed by a discussion of the thermal properties of the earth, and the article concludes with a short section on the origin of the earth's magnetism.

Chapter 4 is by J. Tuzo Wilson (Toronto) who considers " The Development and Structure of the Crust ". It opens with a survey of the general features of the crust, namely ocean basins, continental shelves and coastal plains. Two general causes of these features have been suggested by geophysicists. The first is compression due to thermal contraction of the earth, proposed by Jeffreys in 1929. The second is thermal convection in the mantle, an idea that was first put forward by Vening-Meinesz in 1934. The remainder of Professor Tuzo Wilson's article is devoted to an account of his own extensive development of the thermal contraction hypothesis, which appears to meet with a fair measure of success.

H. V. Sverdrup (Oslo), in his article on " Oceanography " commences with a general description of the ocean basins. He divides the rest of his discussion into two parts, dealing first with the properties of sea water and the complicated heat budget of the oceans, and then with some aspects of dynamical oceanography.

In chapter 6, " Geochemistry of the Crust ", Brian Mason (Indiana University) considers the chemistry of magmatic, sedimentary and metamorphic processes.

The first article on the atmosphere is by H. R. Byers (Chicago), who deals with " The Atmosphere up to 30 Kilometers ". Of the selected topics, viz. the circulation of the troposphere, composition, heat balance and temperature of the region, atmospheric electricity, condensation and precipitation, the first is given most attention.

" Chemical reactions occurring at low levels in the atmosphere are directly or indirectly brought about by living matter at the surface ". These processes are considered in some detail in G. E. Hutchinson's (Yale) chapter entitled " The Biochemistry of the Earth's Atmosphere ", which concludes with a few remarks on the atmospheres of some other planets, notably Mars.

The next five chapters, 9–13, are devoted to the high atmosphere. The constituent gas on the daylight side of this region is partly ionized and excited by incoming solar radiation, and recombination during the night gives rise to feeble emitted light from the night sky. Charged particles emitted by the sun also impinge upon the high atmosphere and, due to influence of the earth's magnetic field, they penetrate most easily at high latitudes on both sides of the earth and cause spectacular current displays.

Leo Goldberg (University of Michigan), who contributed chapter 9, " Absorption Spectrum of the Atmosphere ", deals systematically with the absorption of solar radiation from the ultra-violet to centimetre wavelengths, and with the interpretation of the spectrum in terms of composition and physical state.

The basic data required for the broad characterization of the structure of the high atmosphere are presented by Fred. L. Whipple (Harvard) in his article on " Density, Pressure and Temperature above 30 Kilometres ". Professor Whipple gives the results of very recent measurements obtained with the aid of rockets.

" Emission Spectra of Twilight, Night Sky and Aurorae " are discussed by J. W. Chamberlain and A. B. Meinel (Yerkes and McDonald Observatories) in chapter 11. After a short introduction dealing with the techniques employed, a fairly full account of the results is given. The authors then consider the interpretation of these results, carefully commenting of their significance with respect to existing theories.

Chapter 12 " The Physics of the Upper Atmosphere ", is by D. R. Bates (Belfast) who selects a few topics from a broad field, paying most attention to the photochemistry of the upper atmosphere, to the theory of the formation of the ionized layers, of the airglow and the aurorae, and to the thermal balance. This is followed by M. Nicolet's (Brussels) article on " Dynamic Effects in the High Atmosphere ", which deals mainly with photochemical problems, but concludes with a short account of high atmospheric motions as evinced by photograhic and spectroscopic studies, and by ionospheric and radio-astronomical (scintillation) observations.

The purpose of Clyde T. Holliday's (Johns Hopkins University) short chapter, " The Earth as seen from outside the Atmosphere " is to present eight magnificent photographs taken at different times of year from rockets at approximately 100 km above the S.W. United States.

In the final chapter, " Albedo, Colour and Polarization ", André Danjon (Paris) describes the properties of the earthlight on the moon, from which it is possible to infer something about the appearance of the earth as a planet.

On the whole the book is very well written. Its balance is hard to criticise, since the selection of topics to include in any treatise of this kind is always influenced by many different factors. Perhaps one might have welcomed in this volume an introductory survey of the same type as that given in Volume 1, *The Sun*.

The research workers continuously seek indications of the lines along which future investigations should proceed. Many of these will he find in *The Earth as a Planet*. K. HIDE.

Titanium (*Metallurgy of the Rarer Metals*, Pt. 4), by A. D. McQuillan and M. K. McQuillan. Pp. xix + 466. (London: Butterworths Scientific Publications, 1956.) 56*s*.

This volume, the fourth of the series edited by Dr. H. M. Finniston, on the Metallurgy of the Rarer Metals is larger than its predecessors and is also more expensive. It covers the extraction and fabrication of titanium, its physical properties, the constitution of titanium alloys, the deformation and transformation characteristics of the metal and alloys, mechanical properties and corrosion resistance. The final short chapter is devoted to an account of suitable metallographic techniques. Work in these fields published up to and including 1954 is covered thoroughly and there are occasional references to 1955 papers. The 200 diagrams and photographs are well chosen, though unfortunately the two plates of representative photomicrographs, Figs. 199 and 200, have not come out well. The extraction metallurgy chapter has suffered from excessive condensation, while the description of the fabrication of the metal and alloys might with advantage have been located after the basic physical metallurgy contained in the next six chapters. In a volume of this size some repetition is unavoidable but there is, perhaps, more than is desirable. For example, martensitic transformations are discussed in connection with each alloy system and with considerable duplication in the chapter on transformations. The clear cut distinction drawn between nucleation-and-growth and shear-type transformations in the introduction to this latter chapter will, no doubt, be severely challenged. The myth that most fabrication difficulties with titanium arise from rapid work-hardening is dealt with effectively. Occasionally, lucidity is lost: the statement on page 30 that $TiCl_3$ dissolves in a eutectic mixture of LiCl and KCl but that two immiscible liquids are formed will confuse, so will the statement of the objections to any theory of alloying based upon the extended zone scheme with 4 electrons per atom (page 282). The present work fills a definite need for a readable and reliable reference book and can be recommended to those concerned with the manufacture, fabrication, application, or physical metallurgy of titanium.

P. C. L. PFEIL.

Der elektrische Strom im Gas: Band I, *Allgemeine Probleme der Elektrodynamik der Gase*, by W. L. GRANOWSKI. Pp. xv + 500. (Berlin: Akademie-Verlag, 1955.) DM 44.

The author in his foreword explains that this book is devoted to the physics of the conduction of electricity through gases, theorems of the electrodynamics of gases, fundamentals of the basic theory of elementary processes, and the corresponding experimental data. The book is intended for scientists and engineers and workers in associated fields, scientific and industrial.

The book is in four parts, covering: the electrodynamics of ionized gases; elementary processes such as elastic and inelastic collisions, recombination and the formation of negative ions and phenomena at boundary surfaces; kinetics of electrons and ions in a gas, electron velocity distribution, drift, and ionization; and, finally, a discussion of the general methods of dynamics applied to currents in an ionization gas. There are numerous tables and references.

The treatment is extensive, careful and logical. In much of the mathematical discussion the general methods employed are similar to those in the textbook of Chapman and Cowling. However, a possible criticism is that in view of the amount of space devoted to it one might have expected the theory to be carried further. Similar criticism applies to the account of the experimental work. This edition is dated 1955 but it appears to be a translation of the original Russian work of 1951, and indeed there do not appear to be any references later than this. Consequently, some of the experimental results in important branches are out of date and readers should bear this in mind. It is a pity that so much of the important post-war advances are not incorporated. However, the book can serve as a most full and detailed compilation of the mathematical methods employed in this subject up to 1950. F. LL. J.

Elektrische Vorgänge in Gasen und im Vakuum (Electric phenomena in gases and in a vacuum), by N. A. KAPZOW. (Berlin : Deutscher Verlag der Wissenschaften, 1955.) DM. 42.20.

This is a welcome edition to the series of textbooks on ionization in gases which have appeared in the U.S.S.R. during the post-war period. The subject is apparently regarded there as of considerable importance. The author is the director of a research department of physics in the Lomonossow University at Moscow. The book is designed for students taking advanced courses at Universities and Technical Academies, for industrial researchers and for engineers. The translation into German seems to have been done with great care and the book is on the whole easy to read. It is a tome of 700 pages which differs from our standard textbooks in many respects : it contains not only over 2500 references that deal with various aspects of the movement of charges through a gas, but also sets out to describe the technique of high vacua.

As an introduction the methods of evacuation and of the measurement and preservation of low pressures in vacuum systems are given, as well as the technique of handling gases. Then the thermionic and the field emission processes are treated including the electron emission from mono-layers and oxide cathodes. The Schrot effect, the photoelectric effect, the secondary emission of electrons, the ionization and excitation of gas atoms and the collisions of the second kind are discussed. Then follow the theory of photo and thermal ionization in gases, and sections on negative ions, recombination, mobility, space charge effects and probes. One tenth of the volume is devoted to the emission of radiation from discharges. Other sections contain ionization by collisions in electric fields, the electric breakdown of gases, the glow and the arc discharge, the plasma, the positive column in arcs, the spark discharge, lightning, atmospheric electricity, corona and high frequency discharges and chemical changes in a discharge. The book closes with a short chapter on practical applications.

The text is of great clarity throughout and the presentation breathes modesty and simplicity. Though the classical contributions of foreign authors are on the whole fairly completely given, much emphasis is put on the contributions by Russian and Soviet (*sic !*) scientists whose achievements have in the past probably always been belittled by other nations owing to the language difficulties. On the other hand one finds in various sections of the book, for example in the treatment of secondary processes in breakdown, high frequency discharges, etc., a quite distinct lack of appreciation of modern views which is probably due to the restricted library facilities. If sometimes too much has been attempted and too little depth been reached, this deficiency is balanced by the advantage that the superficial treatment of a great variety of problems might whet the appetite of those readers who may wish to pursue the matter further.

The text is illustrated by 330 figures, many of which can be recognized as old acquaintances, and a great deal of numerical data is given in 67 tables. There is no doubt that those who want to obtain information about many branches of discharge physics will find this book an interesting source of study.

A. VON ENGEL.

Methods of Mathematical Physics, 3rd Edn, by SIR HAROLD JEFFREYS and BERTHA SWIRLES (LADY JEFFREYS). Pp. xii + 714. (Cambridge: University Press, 1956). 84*s*.

When the first edition of this monumental work was issued in 1946 it was fully reviewed in these *Proceedings*, 1947, **59,** 512. Since then a second edition in 1950 corrected a few minor errors, and added some 50 pages of new material. The fact that a third edition is now called for is a tribute to the uniqueness of this book, and the fact that hardly any changes are now required is a tribute to its excellence. The first reviewer said of it that " it is a book of quite exceptional importance ". The present reviewer can merely echo these words. For this account of the methods of mathematical physics is still magnificent in its scope and architecture. There must be very few major techniques in theoretical physics which are not explained here, and illustrated by some genuine example. It remains, and is likely to continue for many years, an almost indispensable work of reference for anyone attempting research in theoretical physics. Even its price, which seems pretty high, is not altogether unreasonable. There will not be many of the people for whom this book is intended who would not save themselves four guineas' worth of time by possessing it.

C. A. COULSON.

Meteors. (Proceedings of a Symposium on Meteor Physics, Jodrell Bank Experimental Station, University of Manchester, July 1954.) Edited by T. R. KAISER. Pp. vii + 204. (London: Pergamon Press, 1955.) 55*s*.

A wide range of investigations are connected with the problems of meteor physics and it was felt that an international conference would provide a valuable meeting for workers in the different fields. Such a conference was held at Jodrell Bank, Manchester, in July 1954 and the articles in this book represent contributions to this symposium.

The book provides, in a convenient form, a survey of the most important current developments in meteor research. The topics which received most attention were the problems of meteor ablation, radio echo studies, the relation between meteors and the ionosphere and the numbers and sizes of meteors in the incident flux. A general measure of agreement was reached over their physical structure and also regarding the effects of meteors on the night-time E region. On the other hand, contributions concerning the relation between radio and photographic meteors showed the need for very much more extensive observations, which can only come from the new fast cameras used in conjunction with the radio measurements. The last part of the book is devoted to accounts of the meteor research programmes in different countries, with particular reference to the International Geophysical Year.

The production and layout are most pleasing and one can only regret the high price, which seems unavoidable in books with a comparatively restricted circulation.

J. E. BALDWIN.

CONTENTS OF SECTION A

PAGE

Dr. W. F. Nash and Mr. A. J. Pointon. The Momentum Spectrum of μ-Mesons at a Depth of 40 Metres Water Equivalent 725

Mr. V. W. Maslen. The Fermi, or Exchange, Hole in Atoms 734

Dr. R. W. Nicholls. The Interpretation of Intensity Distributions in the CN Violet, C_2 Swan, OH Violet and O_2 Schumann–Runge Band Systems by use of their r-Centroids and Franck–Condon Factors 741

Mr. E. E. Maslin, Dr. J. M. Calvert and Dr. A. A. Jaffe. An Investigation of the Reactions: $^{10}B(d, n)^{11}C$ and $^{11}B(d, n)^{12}C$ 754

Mr. T. C. Randle, Dr. D. M. Skyrme, Mr. M. Snowden, Mr. A. E. Taylor, Mr. F. Uridge and Mr. E. Wood. The Scattering of 130 mev Neutrons by Protons 760

Dr. A. C. Hurley. The Binding Energy of the Nitrogen Molecule 767

Prof. C. A. Coulson and Dr. M. J. Stephen. The Quadrupole Polarizability of the Hydrogen Atom 777

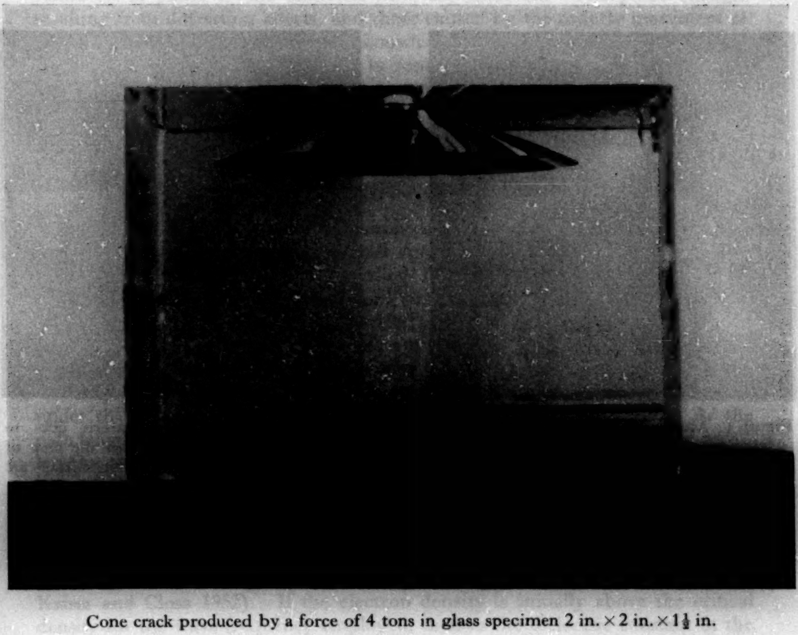
Letters to the Editor:

  Dr. L. F. H. Bovey, Mr. E. B. M. Steers and Mr. H. S. Wise. The Infra-Red Resonance Lines of Lutetium 783

  Dr. A. Dalgarno. Application of the Rayleigh-Schrödinger Perturbation Theory to the Hydrogen Atom 784

Reviews of Books 786

Contents of Section B 788



Cone crack produced by a force of 4 tons in glass specimen 2 in. × 2 in. × 1½ in.

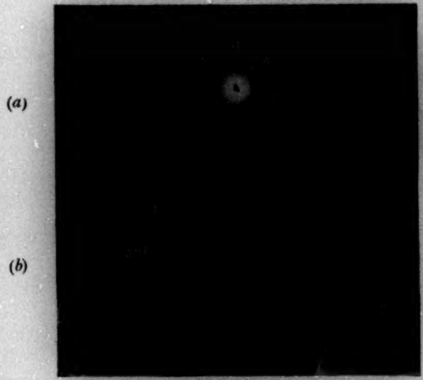


Figure 1. Aluminium oxide pattern obtained from the tongue of oxide shown below.

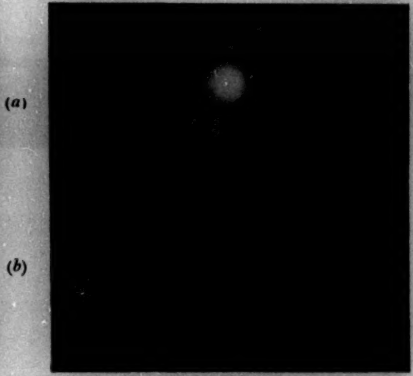


Figure 2. The sharper pattern due to boehmite originating from the area of film shown. Specks of metal are present.

Phase Changes and Resonance Effects in Radio Echoes from Meteor Trails

By J. S. GREENHOW and E. L. NEUFELD
Jodrell Bank Experimental Station, University of Manchester

Communicated by A. C. B. Lovell; MS. received 16*th February* 1956,
and in final form 10*th May* 1956

Abstract. When an ionized meteor trail is studied by the radio echo technique variations in phase of the scattered wave are observed. The phase changes resulting from diffraction effects, and those caused by the orderly movement of the trail under wind forces, are well known. This paper discusses a third type of phase change which is attributed to resonance as the ionized trail expands. It is shown that the measurements are in good agreement with the phase changes predicted by the theory of the scattering of radio waves from meteor trails.

---

§ 1. Introduction

THE changes in phase of the radio echoes from meteor trails due to diffraction effects and the influence of upper atmosphere winds are well understood. The former gives rise to the so called ' Doppler whistle ', and is the basis for the measurement of meteor velocity using continuous wave techniques (Manning, Villard and Peterson 1949, McKinley 1951). The small amplitude fluctuations associated with these diffraction effects are observed with pulsed radar systems, and also give a measure of meteor velocity (Davies and Ellyett 1949).

Large phase changes are also produced by bodily motion of the meteor trails under the influence of winds (Manning, Villard and Peterson 1949). As the path length to and from the echoing point changes by integral numbers of wavelengths, the phase of the echo varies through successive cycles of 360°. Observation of these phase changes enables systematic measurements of the winds between altitiudes of 80–100 km to be made (Greenhow 1954, Elford, Liddy and Robertson 1953).

This paper is concerned with a third type of phase change, which is attributed to resonance effects as a column of meteor ionization diffuses (Herlofson 1951, Kaiser and Closs 1952). If the electron density is initially above the critical density for the wave frequency employed, then as the dielectric constant at the centre of the trail passes through the value $-1{\cdot}4$, the amplitude of the radio echo is enhanced by resonance, provided the electric vector in the incident radio wave is perpendicular to the trail. Such effects have been observed by Closs, Clegg and Kaiser (1953) and Billam and Browne (1955, 1956), who find that the ratio of resonant to non-resonant echo amplitude is approximately 2.

As the meteor trail passes through the resonant condition the theory of scattering predicts that the radio echo should undergo a phase reversal of 180°. The phase angle increases (defined as a positive variation), corresponding to a recession of the echoing point. This paper is concerned with the investigation of phase changes of this type.

§ 2. TECHNIQUE

2.1. *Apparatus*

Changes in echo phase are measured by beating the received signal with a reference waveform of fixed phase. The apparatus has been in use for the past two years for the investigation of upper atmospheric winds, and is fully described elsewhere (Greenhow 1954). Examples of the meteor echoes recorded by the equipment are given in figure 1 (Plate). Various characteristics of the echoes are displayed on three cathode-ray tubes. C.R.T. 1 records the variation of amplitude on a slow time base of 0·3 second duration; the echo pulses are 1/150 second apart. C.R.T. 2 and 3 show the beats in amplitude produced when the coherent echo pulses are combined with various reference waveforms. These signals enable the phase variations between successive echo pulses to be determined as follows.

2.2. *Measurement of Phase Angle*

Let the echo signal received from a meteor trail be represented by

$$y_1 = a(t) \sin [\omega t - \phi(t)]$$

where $a(t)$ is the amplitude of the echo, and $\phi(t)$ its phase relative to some arbitrary zero. Both the amplitude and phase are functions of time. ω is the angular frequency of the radio wave.

When the received signal is added to the reference wave $Y_1 = A_1 \sin \omega t$ of fixed amplitude A_1 and with phase taken as the arbitrary zero, the resultant signal is

$$\begin{aligned} Y_1 + y_1 &= A_1 \sin \omega t + a(t) \sin [\omega t - \phi(t)] \\ &= B_1 \sin (\omega t + \psi_1). \end{aligned}$$

$B_1^2 = A_1^2 + a^2(t) + 2a(t)A_1 \cos \phi(t)$, where B_1 is the amplitude of the combined echo and reference waveforms. For $A_1 \gg a(t)$, this reduces to

$$B_1 = A_1 + a(t) \cos \phi(t).$$

The time varying part of B_1 is the product of echo amplitude and cosine of the phase angle. The quantity $B_1(t) = a(t) \cos \phi(t)$ is displayed on cathode-ray tube 2. This gives the variation of $\cos \phi(t)$ throughout the lifetime of the echo, but it also includes the variable echo amplitude $a(t)$, and does not show whether the phase angle increases or decreases.

These two uncertainties are removed by adding the receiver output to a second reference signal $Y_2 = A_2 \sin (\omega t - \frac{1}{2}\pi)$ in phase quadrature with the first (Manning, Villard and Peterson 1952). Thus

$$Y_2 + y_1 = A_2 \sin (\omega t - \tfrac{1}{2}\pi) + a(t) \sin [\omega t - \phi(t)]$$

and as before this reduces to

$$Y_2 + y_1 = B_2 \sin (\omega t + \psi_2)$$

where $B_2 = A_2 + a(t) \sin \phi(t)$. In this case the time varying part of B_2 is the product of echo amplitude and sine of the phase angle. This quantity $B_2(t)$ is displayed on C.R.T. 3. Now $B_2(t)/B_1(t) = \tan \phi(t)$, thus division of corresponding pulse amplitudes in displays 3 and 2 gives the variation of phase angle throughout the echo, independent of amplitude variations. Inspection of the signs of $\sin \phi(t)$ and $\cos \phi(t)$ gives the sense of the phase change. (In figure 1 the convention is such that if the oscillation in 3 leads that in 2 by $\frac{1}{2}\pi$, the drift of the trail is towards the observer, i.e. the phase angle is decreasing.)

The error in phase measurement is almost entirely limited by the signal-to-noise ratio of the echo. For a ratio of ten times, characteristic of those used in the analysis, the phase angle measured from a single pair of pulses will be in error by approximately 0·14 radian.

§ 3. RESULTS

3.1. *Selection of Echoes*

In general the phase change observed in the radio echo from a meteor trail is predominantly due to upper atmospheric winds. For example in figure 1(*a*) the beat components fluctuate through three cycles (1080°) in 0·26 second. This corresponds to a bodily motion of the trail of 1·5 wavelengths (12·4 m), and a radial component of wind velocity of 48 metres sec^{-1}. Phase changes of this order of magnitude are present in the majority of echoes, and make it difficult to observe the much smaller phase changes due to diffraction and resonance effects. (In any case, because of the low pulse repetition frequency, diffraction fluctuations will not be observed except for very slow meteors.) Thus in the following analysis measurements are restricted to those echoes which show zero or very small phase changes in the main body of the echo.

An example of this type of echo is illustrated in figure 1(*b*). The amplitude variation of $a(t)\sin\phi(t)$ shown in 1(*b*), 3 is almost indistinguishable from the simple amplitude variation $a(t)$ shown in 1(*b*), 1 (although reversed in sign). Thus $\sin\phi(t)$ varies little from -1 throughout most of the echo. In 1(*b*), 2, $\cos\phi(t)$ is such that $a(t)\cos\phi(t)$ is approximately zero for most of the echo, and for this reason the small but rapid variation in $\cos\phi(t)$ near the beginning of the echo is very marked. Indeed division of corresponding pulses between X and Y in 3 and 2 shows that $\tan\phi(t)$ changes from unity to approximately infinity in eight echo pulses. This is a +45° phase change in 0·05 second. (Increase in phase angle as $\cos\phi(t)$ changes from a negative value to zero, while $\sin\phi(t)$ remains negative.)

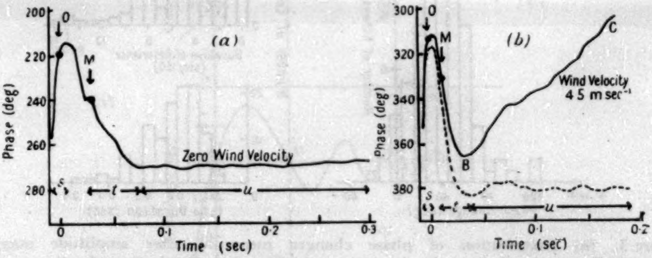


Figure 2. Echo phase ϕ (t) plotted against time measured from passage of meteor across the right-angle reflection point. t, period of resonant phase shift; u phase change caused by bodily motion of the trail; s, phase change due to diffraction effects. O and M mark the position of half maximum and maximum echo amplitude. (*a*) Zero wind. (*b*) Low wind velocity, showing method of correction.

The pulse by pulse variations in phase of this echo are shown in figure 2(*a*). The phase angle first shows a decrease (due to diffraction effects to be discussed later), and this is followed by a rapid increase of 55°. This increase takes place within 50 msec; during the remainder of the echo (0·25 sec) there is no detectable change in phase due to winds. The point of maximum echo amplitude is shown in figure 2(*a*), and the phase change after this time is +30°.

3.2. *Correction for Small Wind Drifts*

An example of an echo showing a small phase change due to bodily motion of the trail is shown in figure 2(*b*). The rapid phase recession OB near the beginning of the echo is quite distinct from the slower variation due to winds in the main body of the echo. In this case the trail is blowing towards the observer. The effect of wind drift is removed by calculating the change of phase per pulse interval from the part of the curve BC and adding this small correction pulse by pulse from A to C. The drift velocity of the trail is approximately 4·5 metres sec^{-1}, or 2·5° per pulse interval. The variation of phase corrected to zero wind velocity is also shown in figure 2(*b*).

In the following analysis echoes are only accepted in which the drift velocity is less than 4 metres sec^{-1} (2° per pulse) as it is desirable to measure resonance phase changes as small as 10° spread over several pulse intervals. A further criterion of course is that echoes are of the short duration 'decay type', with electron line densities of the order of $10^{11}\,\text{cm}^{-1}$. These are the echoes which show resonance effects.

3.3. *Distribution of Observed Phase Changes*

From a total of approximately 35 000 usable echoes obtained during a series of wind measurements (Greenhow and Neufeld 1955 a) only 113 satisfy the above criteria. The remainder of the echoes show wind drifts of 30–40 metres sec^{-1}.†

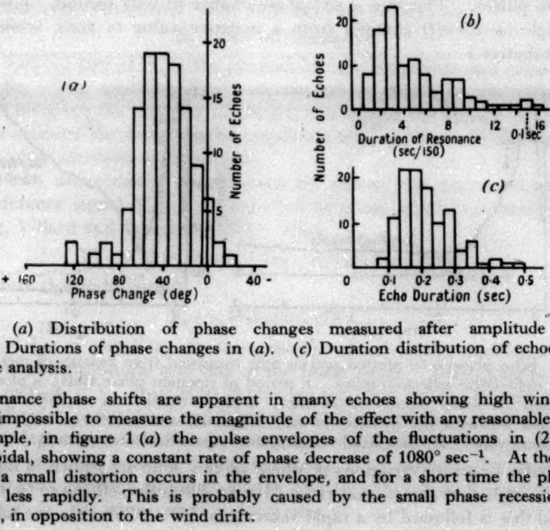


Figure 3. (*a*) Distribution of phase changes measured after amplitude maxima. (*b*) Durations of phase changes in (*a*). (*c*) Duration distribution of echoes used in the analysis.

---

† Resonance phase shifts are apparent in many echoes showing high wind speeds, but it is impossible to measure the magnitude of the effect with any reasonable accuracy. For example, in figure 1 (*a*) the pulse envelopes of the fluctuations in (2) and (3) are sinusoidal, showing a constant rate of phase decrease of 1080° sec^{-1}. At the point *R*, however, a small distortion occurs in the envelope, and for a short time the phase angle decreases less rapidly. This is probably caused by the small phase recession due to resonance, in opposition to the wind drift.

A histogram showing the distribution of phase changes observed after the maximum echo amplitude is given in figure 3(*a*). 92% of the echoes show positive phase shifts lying between 5° and 120°, with a peak in the distribution between 20° and 60°. Only 3% of the echoes show definite negative phase changes.

Figure 3(*b*) is a histogram showing the durations t of the observed phase shifts, measured after the maxima of echo amplitude (figure 2). In 50% of the echoes the phase change is over within 25 msec. The most frequently observed value is 20 msec, which is a small fraction of total echo duration. The distribution of echo durations used in the analysis, measured from time of maximum amplitude to disappearance in receiver noise, is shown in figure 3(*c*). The most probable duration is 0·17 sec. The echo decay times T to $1/e$ of the estimated maximum amplitude for non-resonant scattering are approximately one half those shown in figure 3(*c*). T is related to the wavelength λ and diffusion coefficient D by the expression $T=\lambda^2/16\pi^2 D$ (Herlofson 1948, Greenhow and Neufeld 1955 b).

§ 4. Discussion

4.1. *Phase Changes due to Diffraction*

In the investigation of phase changes in the early part of the radio echo from a meteor trail, diffraction effects during trail formation must be considered. The theoretical amplitude fluctuations are shown in figure 4(*a*), and the phase variations in figure 4(*b*). O corresponds to the point at which the meteor crosses the perpendicular from observer to trail, and at this instant the echo amplitude rises to one half the maximum value at M. Reference to 4(*b*) shows that the

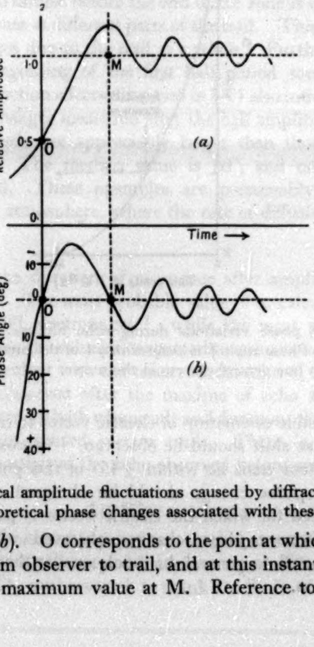


Figure 4. (*a*) Theoretical amplitude fluctuations caused by diffraction effects during trail formation. (*b*) Theoretical phase changes associated with these diffraction effects.

resultant phase change due to diffraction, after the meteor passes the right-angle reflecting point, is zero. During the rapid rise in amplitude before O there is a negative phase change, and this is found in the early parts of echoes shown in figures 2(*a*) and (*b*). After the first maximum of the diffraction pattern the oscillations in phase are only approximately ± 5°. Diffraction effects can therefore be neglected in consideration of the phase changes discussed in this paper, provided measurements are restricted to the time after the passage of the meteor across the right-angle reflecting point. In fact in the analysis the measurements were further restricted to the period after amplitude maxima.

4.2. *Resonance Effects*

When the electric vector in the incident radio wave is perpendicular to the ionized column, theory predicts a phase change of $+180°$ during resonance. According to Kaiser (1955) 135° of the change should occur before maximum, and only 45° after this point (figure 5). When the electric vector is parallel to the trail, of course, no resonance effects are observed. In the series of measurements described in this paper, horizontal polarization was used.

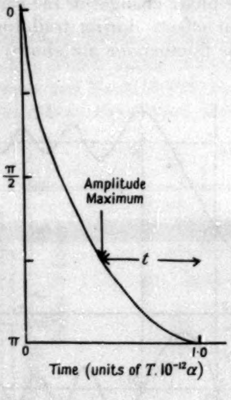


Figure 5. Theoretical phase variations during echo lifetime, for transverse scattering (*after* Kaiser). Phase angle ϕ is defined in § 1 of the present paper, and T in § 3.3. α is the electron line density per cm in the meteor trail.

Thus the most probable orientation of electric vector to trail is 90°, and for such echoes the full phase shift should be observed. Approximately three quarters of the observed meteor trails lie within ± 45° of this position, and more than half the theoretical phase change should be observed even for these echoes. No trails are observed for which the electric vector is parallel to the trail (non-resonant). The resonant phase changes observed with equipments using vertical polarization will, in general, be much smaller than those observed using horizontal polarization.

4.3. *Comparison of Measured Phase Changes with Theory*

(*a*) *Magnitude.*

The phase changes shown in figure 3(*a*) are in good agreement with the theoretical value after the amplitude maxima. The mean value is 38°, compared with the expected figure of 45°. The difference probably arises because of the number of trails observed near parallel polarization. The small number of negative phase changes could arise from minor distortions due to irregular winds. Phase shifts appreciably exceeding the theoretical value are also observed. In the analysis it has been assumed that the amplitude maximum caused by resonance coincides with the diffraction maximum. This is not necessarily the case, as the time of resonance depends upon the rate of diffusion of the ionized column, and upon the initial electron line density. It would be expected, however, that a combination of the two effects would be to displace both the maxima producing a single broad peak, resulting in an uncertainty in determining the exact point from which to measure the phase angle.

Although theory predicts a phase reversal of 180°, changes as large as this are rarely observed. Resonance occurs when the column of ionization has diffused to such a diameter that the dielectric constant at the centre reaches $-1{\cdot}4$. For a trail formed instantaneously the complete resonance would be observed. However, a meteor takes a finite time (of the order of 1/40 sec) to traverse the first Fresnel zone; this is the length of trail primarily responsible for the reflection of radio waves. Thus the ionized matter at the beginning of the Fresnel zone has already started to diffuse before the end of the zone is reached, and resonance occurs at different times at different parts of the trail. This effect tends to smooth out the phase changes due to the trail as a whole. Further, over the section of trail between the beginning of the first half period zone and the right-angle reflecting point, diffraction effects discussed in § 4.1 also influence the observations. Even so the phase changes measured after the half amplitude point on the initial rise in echo amplitude are appreciably larger than those measured after the amplitude maxima. The median value is 60°, and occasionally a full 180° reversal is observed. These examples are presumably due to fast meteors observed low in the atmosphere, where the rate of diffusion is small.

(*b*) *Duration.*

From figure 5 the duration of resonance after amplitude maxima is given by $t \sim 0{\cdot}6T \times 10^{-12\alpha}$. The most probable value of α, calculated from equipment parameters, is 3×10^{11} electrons cm^{-1}, while the most probable value of T is 85 msec (§ 3.3). Thus the value of t expected to occur most frequently is 15 msec. This is in good agreement with the observed duration of 20 msec (§ 3.3). Thus the phase changes observed after the maxima of echo amplitude are in good quantitative agreement, in both magnitude and duration, with theories of scattering of radio waves from ionized meteor trails. There is reasonable qualitative agreement for the portion of the echo observed before amplitude maxima.

Acknowledgments

The work described in this paper was carried out at the Jodrell Bank Experimental Station of the University of Manchester under the direction of Professor A. C. B. Lovell. The authors wish to thank Dr. T. R. Kaiser, Dr. I. C. Browne and Mr. E. R. Billam for stimulating the search for phase changes due to resonance effects, and for their interest in the investigation. One of the authors (J. S. G.) is indebted to Messrs. I.C.I. Ltd. for a Research Fellowship.

REFERENCES

BILLAM, E. R., and BROWNE, I. C., 1955, *J. Atmos. Terr. Phys., Meteors: Special Suppl.*, **2,** 73; 1956, *Proc. Phys. Soc.* B, **69,** 98.

CLOSS, R. L., CLEGG, J. A., and KAISER, T. R., 1953, *Phil. Mag.*, **44,** 313.

DAVIES, J. G., and ELLYETT, C. D., 1949, *Phil. Mag.*, **40,** 614.

ELFORD, W. G., and ROBERTSON, D. S., 1953, *J. Atmos. Terr. Phys.*, **4,** 271,

GREENHOW, J.S., 1954, *Phil. Mag.*, **45,** 364.

GREENHOW, J. S., and NEUFELD, E. L., 1955 a, *Phil. Mag.*, **46,** 376; 1955 b, *J. Atmos. Terr. Phys.*, **6,** 133.

HERLOFSON, N., 1948, *Rep. Progr. Phys.*, **11,** 444 (London: Physical Society); 1951, *Ark. Fys.*, **3,** 247.

KAISER, T. R., 1955, *J. Atmos. Terr. Phys., Meteors: Special Suppl.*, **2,** 73.

KAISER, T. R., and CLOSS, R. L., 1952, *Phil. Mag.*, **43,** 1.

MCKINLEY, D. W. R., 1951, *Astrophys. J.*, **113,** 225.

MANNING, L. A., VILLARD, O. G., and PETERSON, A. M., 1949, *J. Appl. Phys.*, **20,** 475; 1952, *J. Geophys. Res.*, **57,** 387.

Radiative Transfer with Distributed Sources

BY R. G. GIOVANELLI AND J. T. JEFFERIES

Division of Physics, National Standards Laboratory, Commonwealth Scientific and Industrial Research Organization, Sydney

MS. received 21*st December* 1955, *and in revised form* 26*th April* 1956

Abstract. The Eddington approximation is applied to obtain expressions for the total intensity of the diffuse radiation in composite diffusers containing or illuminated by non-uniformly distributed sources. Methods are given for simplifying the solutions for point and line sources in highly scattering media.

---

§ 1. INTRODUCTION

CONSIDERATIONS of radiative transfer were first introduced in astrophysical problems in connection with the formation of absorption lines in solar and stellar spectra. More recently, the theory of radiative transfer has been applied to neutron diffusion and to the optics of diffusing media. The range of problems amenable to exact solution is, however, quite small and for most cases an approximate treatment is the best that can be given. Of the various approximations which have been used for obtaining the total radiation intensity, it seems that the best compromise between consistent success and ease of numerical evaluation is still that introduced by Eddington. This converts an exact integro-differential equation into an approximate second order differential equation and provides a very considerable simplification with generally only a slight loss in accuracy, rarely as large as 5% in those cases where a comparison between exact and approximate results is possible. More elaborate methods are available, however, for many cases where greater accuracy is required.

Attention has so far been largely confined to one-dimensional transfer problems, usually in semi-infinite or plane parallel media although Marshak, Brooks and Hurwitz (1949) have given an exact solution for the mean intensity—corresponding to the neutron density in neutron diffusion theory—in an infinite absorbing and scattering medium illuminated by a uniform point source, and Mark (1944, 1945) has discussed the same problem using spherical harmonics. Composite media of a restricted class have been considered by Berlin (1948), Marshak (1947) and Davison (1951) in connection with the influence on neutron density of completely absorbing spheres or cylinders in non-absorbing infinite media.

Two-dimensional problems have received very little discussion. Some very approximate results have been given by Wallace and Le Caine (1946), while Giovanelli (1956) has given an approximate discussion of the intensity due to a line source within a semi-infinite medium and Jefferies (1955) has given a first approximation to the intensity in a semi-infinite medium illuminated by collimated normally incident radiation the intensity of which varies across the surface. Elliott (1955) has derived asymptotic solutions for a point source on a semi-infinite non-absorbing medium. It seems that progress in the exact solution of such problems will be slow, and in the following we adopt the usual Eddington type approximate diffusion equation in considering rather general cases of non-uniformly distributed sources in composite infinite media containing a plane, cylindrical or spherical discontinuity. Either or both media may be absorbing or non-diffusing.

Methods are presented for simplifying the form of the solutions for point and line sources, thus rendering them more amenable to numerical computation; the techniques are applicable to the solution of a wide range of problems with eccentric point sources.

§ 2. THE DIFFUSION EQUATION AND ITS LIMITATIONS

The approximate equation for the total intensity J of diffuse radiation in a medium of extinction coefficient κ and albedo for single scattering $\varpi_0 = 1 - \lambda$ is given by

$$\nabla^2 J = 3\kappa^2(\lambda J - 4\pi S) \qquad \ldots\ldots(2.1)$$

where

$$J = \int_{4\pi} I d\omega,$$

I being the specific intensity of the diffuse radiation, and S, the source function, is the ratio of the rate of emission of radiation per unit volume and solid angle to the extinction coefficient.

Equation (2.1) contains Eddington's approximation for plane parallel diffusion

$$\int_{-1}^{1} I\mu^2\, d\mu = \tfrac{1}{3} \int_{-1}^{1} I\, d\mu,$$

where $\cos^{-1}\mu$ is the angle to the normal, a condition holding exactly if the intensity can be expanded in a Legendre series

$$I = \sum_{0}^{\infty} c_n \mathrm{P}_n(\mu),$$

provided the coefficient c_2 is zero. In particular, it applies if the radiations in the inward and outward hemispheres are separately isotropic. In media absorbing only slightly, the approximation is very close, but deviations increase as ϖ_0 decreases. Nevertheless, the close agreement between reflectances obtained on this approximation and on exact theory indicates the general suitability of the approximation over a wide range of ϖ_0.

When diffusing media are irradiated non-uniformly from without, the source function S arising from the scattering of the incident radiation is generally an unwieldy complicated integral. A most useful simplification is achieved, however, for a large class of media irradiated diffusely by neglecting the source function and incorporating the incident radiation with the general diffuse radiation field, with suitable boundary conditions. Inside the medium, the intensity of the attenuated incident radiation is far from isotropic except near the surface, $\tau < 0{\cdot}02$ say. If the scattering is fairly high, however, the main component of the radiation field in the interior is the scattered radiation; thus (2.1) applies well to the total radiation field, incident plus scattered, and the source function S may be taken as zero. The range of application for best results is $\varpi_0 \gtrsim 0{\cdot}9$, i.e. for media which in the semi-infinite form have reflectances exceeding about 50%. It applies also to diffusing spheres of small optical thickness for any value of ϖ_0.

§ 3. RANDOM SOURCES IN DISCONTINUOUS MEDIA

In this and the following sections we consider infinite media containing a plane, cylindrical or spherical surface of discontinuity, on either side of which the media have different diffusing properties. As special cases, one of the media may be perfectly transparent or may be black.

The method of solution is very similar for each geometry, and in the present section we restrict discussion to a medium containing a spherical discontinuity of radius r_0. For simplicity, the source is taken to be distributed symmetrically in the azimuth, ϕ, coordinate, but otherwise to be arbitrary, in which case it may be expanded quite generally in the form

$$S(t, \mu) = \sum_{n=0}^{\infty} G_n(t)\,\mathrm{P}_n(\mu) \qquad \ldots\ldots(3.1)$$

where $t = \kappa r\,(3\lambda)^{1/2}$, $\cos^{-1}\mu = \theta$, $G_n(t)$ is a function of t and $\mathrm{P}_n(\mu)$ is the Legendre polynomial of order n.

A solution for the total intensity J may be sought in the form

$$J = \sum_{n=0}^{\infty} \mathcal{J}_n(t)\,\mathrm{P}_n(\mu). \qquad \ldots\ldots(3.2)$$

Substitution of (3.1) and (3.2) in the transfer equation (2.1) then yields

$$\frac{d^2\mathcal{J}_n}{dt^2} + \frac{2}{t}\frac{d\mathcal{J}_n}{dt} - \left[1 + \frac{n(n+1)}{t^2}\right]\mathcal{J}_n = G_n(t) \qquad \ldots\ldots(3.3)$$

whose solution, in the usual notation for Bessel functions of purely imaginary argument, is

$$\mathcal{J}_n(t) = A_n(t)\,t^{-1/2}\,\mathrm{K}_{n+1/2}(t) + B_n(t)\,t^{-1/2}\,\mathrm{I}_{n+1/2}(t). \qquad \ldots\ldots(3.4)$$

Here $A_n(t) = \int_{a_n}^{t} x^{3/2}\,G_n(x)\,\mathrm{I}_{n+1/2}(x)\,dx$, $B_n(t) = \int_{t}^{b_n} x^{3/2}\,G_n(x)\,\mathrm{K}_{n+1/2}(x)\,dx$

and a_n and b_n are constants required to fit the boundary conditions, having different values inside and outside the spherical surface of discontinuity. If there is no source at the centre, $\mathcal{J}_n(t)$ remains finite there, which requires that inside the sphere $a_n = 0$, while, since the total intensity tends to zero at large distances from the source, $b_n = \infty$ outside the sphere. The other two sets of constants are obtained by equating total intensities and radial fluxes on either side of the spherical surface. Thus at $r = r_0$,

$$J_1 = J_2, \quad \frac{1}{\kappa_1}\frac{\partial J_1}{\partial r} = \frac{1}{\kappa_2}\frac{\partial J_2}{\partial r} \qquad \ldots\ldots(3.5)$$

the left- and right-hand sides of (3.5) referring respectively to the interior and exterior regions. The solution of (3.4) subject to (3.5) is lengthy but straightforward.

In the special case of a sphere in a vacuum, the second boundary condition (3.5) may be approximated by

$$J + \frac{2}{3\beta}\frac{\partial J}{\partial t} = 0 \qquad \ldots\ldots(3.6)$$

where $\beta = (3\lambda)^{-1/2}$, which leads to the solution for the nth component of J,

$$\mathcal{J}_n(t) = t^{-1/2}\,\mathrm{K}_{n+1/2}(t)\int_0^t x^{3/2}\,G_n(x)\,\mathrm{I}_{n+1/2}(x)\,dx + t^{-1/2}\,\mathrm{I}_{n+1/2}(t)\left[\int_t^{t_0} x^{3/2}\,G_n(x)\,\mathrm{K}_{n+1/2}(x)\,dx + X_n\right] \qquad \ldots\ldots(3.7)$$

where

$$X_n = \frac{t_0^{-1/2} \mathrm{K}_{n+1/2}(t_0) + \frac{2}{3}\beta^{-1}[d/dt\{t^{-1/2}\,\mathrm{K}_{n+1/2}(t)\}_{t=t_0}]}{t_0^{-1/2}\,\mathrm{I}_{n+1/2}(t_0) + \frac{2}{3}\beta^{-1}[d/dt\{t^{-1/2}\,\mathrm{I}_{n+1/2}(t)\}_{t=t_0}]} \int_0^{t_0} x^{3/2}\, G_n(x)\, \mathrm{I}_{n+1/2}(x)\, dx.$$

In the case of a spherical hole in an infinite diffuser, the net radial flux is zero at the surface, whence dJ/dt there is zero. For a black sphere in an infinite diffuser, the flux out from the sphere is zero, and

$$J - \frac{2}{3\beta}\frac{\partial J}{\partial t} = 0$$

on the sphere surface. The corresponding solutions follow readily.

§ 4. MEDIA IRRADIATED DIFFUSELY

When an isolated semi-infinite medium, cylinder or sphere, $\varpi_0 \gtrsim 0{\cdot}9$, is irradiated diffusely from the outside, the results of § 3 can be greatly simplified by neglecting the source function and introducing appropriate boundary conditions. Solutions are given below for arbitrary distributions of the incident radiation in one dimension only. While the results are of intrinsic interest, their main application is in the study of diffusion from point and line sources, discussed in the next section.

4.1. *Semi-infinite Medium*

Let the surface of a semi-infinite diffuser be irradiated isotropically, the incident flux per unit area $F_0(x)$ being a function of the x coordinate only. If the z axis is normal to the surface, then when $\varpi_0 \gtrsim 0{\cdot}9$, (2.1) may be replaced by

$$\frac{\partial^2 J}{\partial u^2} + \frac{\partial^2 J}{\partial w^2} = J \qquad \text{......(4.1.1)}$$

where J is the sum of the total intensities of the incident radiation and of the scattered beam, and $u = \kappa x(3\lambda)^{1/2}$, $w = \kappa z(3\lambda)^{1/2}$. The solution of (4.1.1) is required subject to the boundary conditions

$$J - 2F = 4F_0(x) \qquad \text{......(4.1.2)}$$

where

$$F = \int_{4\pi} I\mu\, d\omega,$$

$\cos^{-1}\mu$ being the angle between a ray and the z axis.

A formal solution of (4.1.1) may be found using a Fourier transform method. The transform of (4.1.1) with respect to u is

$$\frac{d^2\mathscr{J}}{dw^2} = (1+\xi^2)\mathscr{J} \qquad \text{......(4.1.3)}$$

where

$$\mathscr{J} = (2\pi)^{-1/2}\int_{-\infty}^{+\infty} J \exp(i\xi u)\, du.$$

The solution of (4.1.3) which does not diverge as $z \to -\infty$ is

$$\mathscr{J} = A \exp[w(1+\xi^2)^{1/2}] \qquad \text{......(4.1.4)}$$

and the arbitrary constant A follows from the transform of (4.1.2), noting that $F = -\frac{1}{3}\beta^{-1}\,\partial J/\partial w$. Thus

$$A = \frac{4\psi}{1 + \frac{2}{3}\beta^{-1}(1+\xi^2)^{1/2}} \qquad \text{......(4.1.5)}$$

where $$\psi=(2\pi)^{-1/2}\int_{-\infty}^{+\infty} F_0(u)\exp(i\xi u)\,du.$$

Hence $$J=(2\pi)^{-1/2}\int_{-\infty}^{+\infty}\frac{4\psi\exp(-i\xi u)\exp[w(1+\xi^2)^{1/2}]\,d\xi}{1+\frac{2}{3}\beta^{-1}(1+\xi^2)^{1/2}}. \qquad \ldots\ldots(4.1.6)$$

In particular, for a strip source of infinitesimal width lying in the surface and emitting total flux $2\pi Q$ per unit length, $F_0(u)=2\pi Q\delta(u)$ where $\delta(u)$ is Dirac's delta function, and

$$J=2\pi Q\int_{-\infty}^{+\infty}\frac{\exp(-i\xi u)\exp[w(1+\xi^2)^{1/2}]\,d\xi}{1+\frac{2}{3}\beta^{-1}(1+\xi^2)^{1/2}}$$
$$=4\pi Q\int_{0}^{\infty}\frac{\cos(\xi u)\exp[w(1+\xi^2)^{1/2}]\,d\xi}{1+\frac{2}{3}\beta^{-1}(1+\xi^2)^{1/2}}. \qquad \ldots\ldots(4.1.7)$$

If the irradiation over the surface is periodic so that

$$4F_0(u)=\sum_{n=0}^{\infty}\alpha_n\cos nu \qquad (n=0,\ 2\pi k,\ 4\pi k,\ \ldots)$$

the corresponding solution is

$$J=\sum_{n=0}^{\infty}\frac{\alpha_n\exp[w(1+n^2)^{1/2}]\cos nu}{1+\frac{2}{3}\beta^{-1}(1+n^2)^{1/2}}. \qquad \ldots\ldots(4.1.8)$$

4.2. *Cylindrical Medium*

Let the surface of an infinitely long cylinder of radius ρ_0 be irradiated isotropically, the incident flux per unit area $F_0(\phi)$ being a function of ϕ only. Then the solution corresponding to (4.1.8) is

$$J=\sum_{n=0}^{\infty}\frac{\alpha_n\cos(n\phi)\,\mathrm{I}_n(t)}{\mathrm{I}_n(t_0)+\frac{2}{3}\beta^{-1}[\mathrm{I}_n'(t)]_{t=t_0}} \qquad \ldots\ldots(4.2.1)$$

where $$4F_0(\phi)=\sum_{n=0}^{\infty}\alpha_n\cos n\phi,$$

$t=\kappa\rho(3\lambda)^{1/2}$ and $\mathrm{I}_n(t)$ is the modified Bessel function of purely imaginary argument.

4.3. *Spherical Medium*

Let the surface of a sphere of radius r_0 be irradiated isotropically, the incident flux per unit area $F_0(\mu)$ being a function of θ only. Then the total intensity at any point is

$$J=\sum_{n=0}^{\infty}\frac{\alpha_n t^{-1/2}\,\mathrm{I}_{n+1/2}(t)\,\mathrm{P}_n(\mu)}{t_0^{-1/2}\,\mathrm{I}_{n+1/2}(t_0)+\frac{2}{3}\beta^{-1}(d/dt)[t^{-1/2}\,\mathrm{I}_{n+1/2}(t)]_{t=t_0}} \qquad \ldots(4.3.1)$$

which may be written as

$$J=\sum B_n t^{-1/2}\,\mathrm{I}_{n+1/2}(t)\,\mathrm{P}_n(\mu) \qquad \ldots\ldots(4.3.2)$$

where $t=\kappa r(3\lambda)^{1/2}$ and $4F_0(\mu)=\sum_{n=0}^{\infty}\alpha_n\mathrm{P}_n(\mu)$.

§ 5. ECCENTRIC POINT AND LINE SOURCES

The formal solution of the type (3.7) meets with some difficulties in the case of an eccentric point source, for it is then necessary to expand the source function in terms of Legendre polynomials of high degree. The solutions thus require lengthy numerical computation in any particular case.

For high scattering, $\varpi_0 \gtrsim 0{\cdot}9$, simpler solutions generally requiring only a small number of terms for adequate evaluation can be obtained for a medium containing a spherical discontinuity and an eccentric point source, analogous expressions being obtained for other geometries. The technique is demonstrated below in discussing diffusion in a semi-infinite medium and, in greater detail, an infinite medium containing a spherical discontinuity.

5.1. *Line Source Parallel to a Semi-infinite Diffuser*

Consider a line source emitting flux $4\pi Q$ per unit length uniformly into a semi-infinite medium, the source being parallel to the surface. At distance ρ from the source, let the total intensity be $QM_1(\rho)$.

Now suppose a second identical semi-infinite medium placed against the first, the total intensity being increased to $Q\mathscr{F}_1(\rho)$, its magnitude in an infinite medium. $\mathscr{F}_1(\rho)$ has been discussed and evaluated by Jefferies (1955) using an Eddington approximation.

Using a method introduced by Giovanelli (1955), it is now noted that $Q\mathscr{F}_1(\rho)$ consists of two components, the first of which $QG_1(\rho)$ is due to radiation which has at all times been confined to the first semi-infinite diffuser. The second is due to radiation which has diffused once or more through the adjacent medium, resulting in a diffuse flux of magnitude

$$F_0 = \tfrac{1}{4}\left\{J + \frac{2}{3\kappa}\frac{\partial J}{\partial z}\right\}_{z=0} = \tfrac{1}{4}Q\left\{\mathscr{F}_1(\rho') + \frac{2}{3\kappa}\frac{\partial}{\partial z}\mathscr{F}_1(\rho')\right\}$$

entering the first diffuser at all points of its surface, the distance ρ' being that of a surface point from the source.

Since the intensity along the surface is closely isotropic, equation (4.1.6) gives the value J_{P} of the second component. Then

$$QM_1(\rho) = Q\mathscr{F}_1(\rho) - J_{\mathrm{P}}. \qquad \ldots\ldots(5.1.1)$$

For the total intensity at the surface due to a line source on the surface, J_{P} simplifies to

$$J_{\mathrm{P}} = \left(\frac{2}{\pi}\right)^{1/2}\int_0^\infty \frac{4\psi \cos \xi u \, d\xi}{1 + \frac{2}{3}\beta^{-1}(1+\xi^2)^{1/2}} \qquad \ldots\ldots(5.1.2)$$

where

$$4\psi = Q\left(\frac{2}{\pi}\right)^{1/2}\int_0^\infty \mathscr{F}_1(u)\cos \xi u \, du.$$

5.2. *Isolated Sphere*

Consider a uniform point source S emitting flux $4\pi Q$ at distance z_0 from the centre of a diffusing sphere of radius r_0. At an arbitrary point, distance z from the source, let the total intensity be $QM_2(z)$. Now suppose the sphere to be surrounded by a diffuser of the same properties as itself. The source is then in an infinite diffuser, the total intensity at the above point being $Q\mathscr{F}_2(z)$. The function $\mathscr{F}_2(z)$ has been described and evaluated by Jefferies (1956) using an Eddington approximation.

It is now noted that $Q\mathscr{F}_2(z)$ consists of two components, the first of which $QM_2(z)$ is due to radiation which has at all times been confined to the sphere. The second is due to radiation which has diffused once or more through the outer region. There is thus a radial flux entering the sphere, from all points on its surface, whose magnitude is

$$F_0 = \tfrac{1}{4}\left[J + \frac{2}{3\kappa}\frac{\partial J}{\partial r}\right]_{r=r_0} = \frac{Q}{4}\left[\mathscr{F}_2(z) + \frac{2}{3\kappa}\frac{\partial \mathscr{F}_2(z)}{\partial r}\right]_{r=r_0}. \qquad \ldots\ldots(5.2.1)$$

The results of § 4.3 then give as the total intensity due to this component $Q\Sigma\mathcal{J}_n(t)\,\mathrm{P}_n(\mu)$ where

$$Q\mathcal{J}_n(t) = B_n t^{-1/2}\,\mathrm{I}_{n+1/2}(t) \qquad \ldots\ldots(5.2.2)$$

and B_n is as in (4.3.2) with

$$4F_0 = Q\left[\mathcal{F}_2(z) + \frac{2}{3\kappa}\frac{\partial \mathcal{F}_2(z)}{\partial r}\right]_{r=r_0}. \qquad \ldots\ldots(5.2.3)$$

Thus

$$QM_2(z) = Q\mathcal{F}_2(z) - \sum_{n=0}^{\infty} B_n t^{-1/2}\,\mathrm{I}_{n+1/2}(t)\,\mathrm{P}_n(\mu).$$

5.3. *Point Source in a Sphere in an Infinite Diffuser*

Suppose the sphere of the previous example is immersed in an infinite diffusing medium of different properties, the source being far enough within the sphere for direct irradiation of the outer medium to be negligible. Then the total intensity in the sphere due to radiation which has not crossed the boundary at least once is $QM_2(z)$.

To obtain the contribution due to radiation which has traversed the outer diffusing medium at least once, suppose first that the outer diffusing medium is removed. Then on the boundary of the sphere at a distance y from the source,

$$QM_2(y) = Q\mathcal{F}_2(y) - \sum_{n=0}^{\infty} B_n t_0^{-1/2}\,\mathrm{I}_{n+1/2}(t_0)\,\mathrm{P}_n(\mu). \qquad \ldots\ldots(5.3.1)$$

As this radiation is all directed outwards, the outward flux with a vacuum outside the sphere would be $F_0 = \frac{1}{2}QM_2(y)$ per unit area. However, with a diffusing medium outside, this flux may be scattered, and diffuses through both media. For this component, the equation of transfer for high scattering is then $\nabla^2 J = 3\kappa^2\lambda J$ with solutions of the form

$$J_1 = Q\sum_{n=0}^{\infty}\mathcal{J}_{1,n}(t)\,\mathrm{P}_n(\mu), \qquad J_2 = Q\sum_{n=0}^{\infty}\mathcal{J}_{2,n}(s)\,\mathrm{P}_n(\mu)$$

where subscripts 1 and 2 refer respectively to the inner and outer media and $t = (3\lambda_1)^{1/2}\kappa_1$, $s = (3\lambda_2)^{1/2}\kappa_2$. With the boundary conditions that J_1 is finite at $t = 0$ and J_2 is zero at $s = \infty$,

$$\mathcal{J}_{1,n} = C_n t^{-1/2}\,\mathrm{I}_{n+1/2}(t), \quad \mathcal{J}_{2,n} = D_n s^{-1/2}\,\mathrm{K}_{n+1/2}(s).$$

The constants C_n and D_n are obtained from the boundary conditions at $r = r_0$ where the intensities J and radial fluxes F change discontinuously:

$$\left.\begin{aligned} &[J_2 - J_1]_{r=r_0} = QM_2(y) \\ &[F_2 - F_1]_{r=r_0} = \tfrac{1}{3}\left[\kappa_1^{-1}\frac{\partial J_1}{\partial r} - \kappa_2^{-1}\frac{\partial J_2}{\partial r}\right]_{r=r_0} = \tfrac{1}{2}QM_2(y). \end{aligned}\right\} \qquad \ldots\ldots(5.3.2)$$

Expanding $M_2(y)$ in the form $\sum_{n=0}^{\infty} m_n\,\mathrm{P}_n(\mu)$ where

$$m_n = (n + \tfrac{1}{2})\int_{-1}^{+1} M_2(y)\,\mathrm{P}_n(\mu)\,d\mu$$

we find equations for C_n and D_n of the form

$$\left.\begin{aligned} &D_n\,\mathrm{K}_{n+1/2}(s_0) - C_n\,\mathrm{I}_{n+1/2}(t_0) = m_n t_0^{1/2} \\ &C_n\beta_1^{-1}\frac{d}{dt}[t^{-1/2}\,\mathrm{I}_{n+1/2}(t)]_{r=r_0} - D_n\beta_2^{-1}\frac{d}{ds}[s^{-1/2}\mathrm{K}_{n+1/2}(s)]_{r=r_0} = \tfrac{3}{2}m_n. \end{aligned}\right\} \qquad \ldots(5.3.3)$$

It follows that the total intensity inside the sphere is

$$J_1 = QM_2(z) + Q\sum_{n=0}^{\infty} C_n t^{-1/2}\,\mathrm{I}_{n+1/2}(t)\,\mathrm{P}_n(\mu)$$

and outside

$$J_2 = Q\sum_{n=0}^{\infty} D_n s^{-1/2}\,\mathrm{K}_{n+1/2}(s)\,\mathrm{P}_n(\mu). \qquad \ldots\ldots(5.3.4)$$

5.4. *Black Sphere in an Infinite Diffuser*

Provided that negligible direct radiation from the source would be transmitted through the further side of the spherical discontinuity if it were replaced by a material having the same properties as the external medium, the diffusion of radiation around a black sphere can be treated as in § (5.1). At an arbitrary point in the outer medium distant z from a uniform point source emitting flux $4\pi Q$ and also situated in the outer medium, let the total intensity be $QM_3(z)$. Then it is found that

$$M_3(z) = \mathscr{F}_2(z) - \sum_{n=0}^{\infty} A_n t^{-1/2}\,\mathrm{K}_{n+1/2}(t)\,\mathrm{P}_n(\mu) \qquad \ldots\ldots(5.4.1)$$

where

$$A_n\left\{t_0^{-1/2}\,\mathrm{K}_{n+1/2}(t_0) - \tfrac{2}{3}\beta^{-1}\frac{d}{dt}\left[t^{-1/2}\,\mathrm{K}_{n+1/2}(t)\right]_{t=t_0}\right\}$$

$$= (n+\tfrac{1}{2})\int_{-1}^{+1}\left[\mathscr{F}_2(z) - \tfrac{2}{3}\beta^{-1}\frac{\partial \mathscr{F}_2(z)}{\partial t}\right]_{r=r_0} \mathrm{P}_n(\mu)\,d\mu.$$

5.5. *Point Source in an Infinite Diffuser Containing a Diffusing Sphere*

Provided the source is far enough away for direct irradiation of the sphere to be negligible, the total intensity can be found as in § 5.2. Inside the sphere, the total intensity is, in the notation of § 5.4,

$$J_1 = Q\sum_{n=0}^{\infty} C_n' t^{-1/2}\,\mathrm{I}_{n+1/2}(t)\,\mathrm{P}_n(\mu)$$

while outside

$$J = QM_3(z) + Q\sum_{n=0}^{\infty} D_n' s^{-1/2}\,\mathrm{K}_{n+1/2}(s)\,\mathrm{P}_n(\mu)$$

where C_n' and D_n' are given by the equations

$$C_n' t_0^{-1/2}\,\mathrm{I}_{n+1/2}(t_0) - D_n' s_0^{-1/2}\,\mathrm{K}_{n+1/2}(s_0) = g_n$$

$$C_n'\beta_1^{-1}\frac{d}{dt}\left[t^{-1/2}\,\mathrm{I}_{n+1/2}(t)\right]_{t=t_0} - D_n'\beta_2^{-1}\frac{d}{ds}\left[s^{-1/2}\,\mathrm{K}_{n+1/2}(s)\right]_{s=s_0} = \tfrac{3}{2}g_n$$

and

$$g_n = (n+\tfrac{1}{2})\int_{-1}^{+1} M_3(y)\,\mathrm{P}_n(\mu)\,d\mu.$$

The case of a point source on the surface of a sphere can also be handled by the above technique provided negligible radiation passes directly through the sphere.

REFERENCES

BERLIN, T. H., 1948, *Phys. Rev.*, **73**, 437.

DAVISON, B., 1951, *Proc. Phys. Soc.* A, **64**, 881.

ELLIOTT, J. P., 1955, *Proc. Roy. Soc.* A, **228**, 424.

GIOVANELLI, R. G., 1956, *Optica Acta*, **3**, 24, 49.

JEFFERIES, J. T., 1955, *Optica Acta*, **2**, 109, 163.

MARK, J. C., 1944, *N.R.C. Atomic Energy Project*, publication No. NRC 1588; 1945, *Ibid.*, No. NRC 1589.

MARSHAK, R. E., 1947, *Phys. Rev.*, **71**, 443, 688.

MARSHAK, R. E., BROOKS, H., and HURWITZ, H., 1949, *Nucleonics*, **4**, No. 5, 10.

WALLACE, P. R., and LE CAINE, J., 1946, *N.R.C. Atomic Energy Project*, publication No. NRC 1480.

Optical Absorption in Liquid Sulphur–Selenium Co-polymers

By R. A. HYMAN
Standard Telecommunication Laboratories Ltd., Enfield, Middlesex

Communicated by R. W. Ditchburn; *MS. received 28th May* 1956

Abstract. The position of the optical absorption edge in a range of sulphur–selenium co-polymers has been determined as a function of temperature and composition. The temperature dependence of the absorption edge is $-1{\cdot}8 \times 10^{-3}$ ev deg^{-1} in selenium and co-polymers containing up to 84 atomic % sulphur; and $-2{\cdot}9 \times 10^{-3}$ ev deg^{-1} in sulphur.

§ 1. Introduction

Sulphur and selenium each occur in several chain forms (cf. von Hippel 1948); each atom is bound to its two neighbours in the chain by single covalent bonds, and the chains are bound together by residual forces. In particular the chains may be bent to form closed rings.

Two crystalline forms of sulphur have been definitely identified, both based on ring-of-eight molecules, and having respectively rhombic and monoclinic symmetries. There is also an amorphous form, composed of long chains randomly placed, which is obtained by quenching the liquid from a high temperature. The liquid itself has two phases: it is normally fluid below 160°c, composed mainly of ring-molecules; these break at temperatures above 160°c and polymerize to form chains, making the liquid viscous.

Selenium has a monoclinic and an amorphous form corresponding to forms of sulphur. In addition it has a hexagonal form based on long spiral chains lying parallel to each other. The liquid corresponds to the viscous phase of liquid sulphur.

Sulphur and selenium are miscible in all proportions and a range of mixed crystals occurs.

In general it may be noted that the optical absorption edge is similar in the various chain-structures of sulphur and selenium respectively, in spite of their differing overall structures. This suggests that the optical absorption edge is, to a first approximation, determined by the intramolecular structure, that is, simply by the bonding along the chains.

Using samples of about 40 mm thickness, some workers (Mondain-Monval *et al.* 1930) found that the absorption edge in liquid sulphur is displaced towards the red by 2·94 Å deg^{-1} up to 160°c, and by 7·4 Å deg^{-1} at higher temperatures. With thin films, more recent work (Bass 1953) shows a displacement of 6·20 Å deg^{-1} towards the red, both above and below 160°c. The absorption edge in liquid selenium is also displaced towards longer wavelengths, by $1{\cdot}4 \times 10^{-3}$ ev deg^{-1} (Saker 1952). The present work has been previously partly discussed (Hyman 1955).

§ 2. Experimental Method

Pure sulphur and selenium in the form of fine powders were mixed together and then fused, kept nearly at the boiling point for six hours and finally cast into rods.

Three fused quartz absorption cells were used, having path lengths of 5, 10 and 20 mm respectively. A constant deviation rocksalt monochromator was used which had a bandwidth of approximately 0·01 μ at 0·7 μ, and of approximately 0·03 μ at 1·3 μ. The light was mechanically chopped and the detector used was a lead sulphide photo conductive cell in conjunction with a sensitive tuned amplifier (Brown 1952). All measurements were made when the temperature had previously been reduced, but in each case the cell was kept at constant temperature for fifteen minutes before measurements were made. With this precaution there was, however, no evidence that the direction of previous temperature change affected the results.

The alloy containing 59 atomic % of sulphur was reheated and brought to the boil for a further three hours, but no change in the absorption was observed, indicating that the alloy had been well mixed.

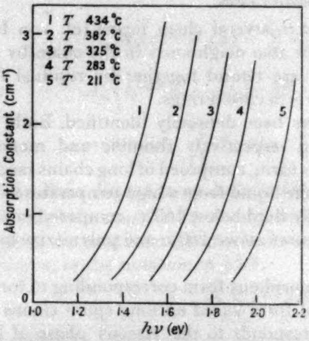


Figure 1. Absorption constant of liquid sulphur.

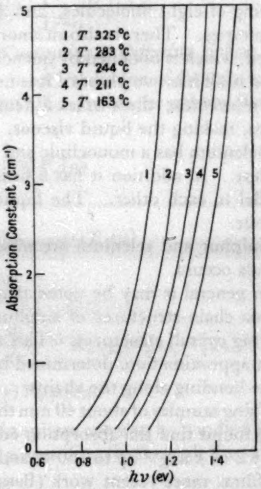


Figure 3. Absorption constant of liquid alloy 24%S–76%Se.

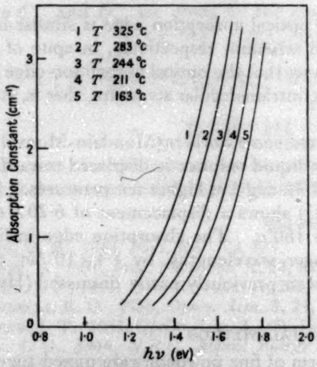


Figure 2. Absorption constant of liquid alloy 84%S–16%Se.

§ 3. Experimental Results

Figures 1, 2 and 3 show, respectively, the absorption curves of sulphur, a co-polymer of 84 atomic % sulphur–16 atomic % selenium, and a co-polymer of 24 atomic % sulphur – 76 atomic % selenium. In each case there exists an absorption edge, that is, a spectral range in which the absorption constant rises steeply from a low value, less than $0{\cdot}1\ \mathrm{cm}^{-1}$ on the low frequency side, to a value too high to be measured with the apparatus used in the present experiment and probably greater than $10^4\ \mathrm{cm}^{-1}$ on the high frequency side. Since the slope of the curve is finite, it is necessary to choose arbitrarily some value of the absorption constant, $\mu = 2\ \mathrm{cm}^{-1}$ in this paper, to define the edge. We write $E_{\mu=2}$ for the quantum energy of the absorption edge defined in this way.

In selenium the edge is displaced by $1{\cdot}8 \times 10^{-3}$ ev deg^{-1} and in sulphur by $2{\cdot}9 \times 10^{-3}$ ev deg^{-1}, each over a range of temperatures of 211–434°c. In the co-polymers with up to 84 atomic % sulphur, and measured over a range of temperatures of 163–325°c, the rate of displacement of the absorption edge as a function of temperature is equal to that of selenium.

§ 4. Discussion

If the absorption in the alloys were simply an additive function of the absorption in the elements taken separately, the edge in the co-polymers investigated would remain nearly in the position of the edge of selenium. Figure 4 shows that this is not so. On the other hand, the temperature dependence of the edge in the alloys is nearly equal to that in selenium. We may give a qualitative picture of these observations as follows:

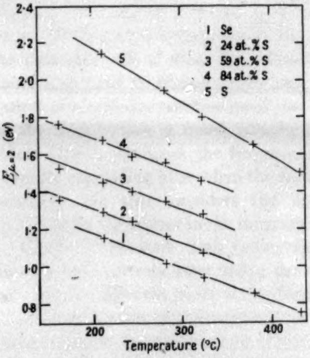


Figure 4. Temperature dependence of absorption edge ($\mu = 2\ \mathrm{cm}^{-1}$) in liquid Se–S alloys.

Suppose that in the co-polymers the selenium and sulphur are completely mixed on an atomic scale, and that the absorption is determined by regions—perhaps lengths of chain—then, provided the regions are not too large, a random distribution will include regions in which there is a large preponderance of selenium, which will absorb as pure selenium and will be mainly responsible for absorption at low values of the absorption constant, that is, at lower frequencies. The absorption reaches the value $\mu = 2\ \mathrm{cm}^{-1}$ only at a frequency at which pure selenium would have a much higher value of absorption constant. The frequency at which this happens will increase as the number of selenium dominated regions decreases so that the absorption edge moves continuously to higher frequencies as the proportion of selenium is reduced.

ACKNOWLEDGMENTS

This work was carried out at Reading University. The author is much indebted to Professor R. W. Ditchburn for advice and the provision of facilities for research and to Standard Telecommunication Laboratories for leave of absence and support.

REFERENCES

BASS, A. M., 1953, *J. Chem. Phys.*, **21,** 80.
BROWN, D. A. H., 1952, *J. Sci. Instrum.*, **29,** 292.
VON HIPPEL, A., 1948, *J. Chem. Phys.*, **16,** 372.
HYMAN, R. A., 1955, *Nuovo Cim.*, Suppl. **2,** series X, 754.
MONDAIN-MONVAL, P., JOB, R., and GALLET, P., 1930, *Bull. Soc. Chim. Franc.*, **47,** 545.
SAKER, E. W., 1952, *Proc. Phys. Soc.* B, **65,** 785.

Spatial Distribution of Fragments of Explosively Loaded Thin-Walled Steel Cylinders

By SAMPOORAN SINGH

Defence Science Laboratory, Ministry of Defence, New Delhi, India

Communicated by F. C. Auluck; *MS. received* 22*nd May* 1956

Abstract. This paper describes the spatial distribution of fragments on static detonation of a thin-walled steel cylinder internally loaded with explosives. The fragments on striking a mild steel 'witness-plate' produce a fragmentation pattern and a simple analysis of its geometry indicates the direction of emission of the fragments. Experimental results indicate that the direction of emission of most of the fragments make an angle of 7·5° to the normal to the inner surface of the cylinder and away from the exploder.

---

§ 1. INTRODUCTION

WHEN a detonating explosive in a metallic cylinder (or shell) is fired at one end, the detonation wave moves down the cylinder and the explosive turns into gaseous products which exert extremely high pressures on the inside of the cylinder. The initial pressure is of the order of 0·3 million atmospheres (Pack *et al.* 1948) and the duration of the pressure is only a few microseconds. If the cylinder wall is thin, the cylinder will expand radially under the high pressure and tend to fracture into long thin fragments. The problem of fragmentation of a cylinder resolves itself into finding the mechanism of fragmentation, and the distribution in mass, velocity and spatial distribution of the fragments. Mott (1947) considers the fragmentation of a shell case as the tearing apart of a rapidly expanding tube when the material of the tube reaches the limits of its ductility. He also considers the distribution of fragment dimensions to be determined by the scatter in the magnitude of the strain required to produce fracture. Clark (1949) took flash radiographs of the 20 mm high explosive shell during and immediately after static detonation and determined the initial fragment velocity for different parts of the fragmentation pattern.

The present paper deals with some experimental work on the spatial (angular) distribution of fragments on static detonation of an explosively loaded thin-walled steel cylinder. The fragments strike a mild steel 'witness-plate' and record on it a fragmentation pattern. A simple analysis of its geometry indicates the direction of emission of the fragments.

§ 2. EXPERIMENTAL WORK

2.1. Cylinders were machined from solid round commercial steel bars and annealed or cut from 'water quality' pipe. The inside diameter of each cylinder was in the range between 1½ in. to 2½ in. and wall thicknesses varied from ⅛ in. to ⅜ in. All the cylinders were made 5 in. long. The cylinder was mounted in a horizontal position having its axis at a certain convenient height from a mild steel 'witness-plate'. The outer surface of each cylinder was free from any restraint. High explosive (60 RDX, 40 TNT) was cast in the equipment. The charge was boosted by a 1 oz guncotton primer and initiated by an A.S.A./Tetryl electric detonator. The lay-out of the equipment for firing is shown in figure 1 (*a*). The particular cylinders that were fired are listed in table 1.

Table 1. Test Cylinder Data

Material	Water-quality pipe		Machined from bars			
Inner diameter (in.)	$1\frac{1}{2}$	2	$2\frac{1}{2}$	2	2	2
Wall thickness (in.)	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{5}{16}$	$\frac{3}{8}$
No. of cylinders fired	4	8	4	5	4	3

On firing, the cylinder broke into fragments which struck the witness-plate and produced on it a fragmentation pattern. A typical fragmentation pattern is shown in figure 1 (*b*). This consists of (*a*) scooping of the metal of the witness-plate in a direction parallel to the axis of the cylinder due to the impact of fragments arising from the main body of the cylinder on the witness-plate, and (*b*) an arc RS of a circle due to the impact of fragments arising from the forward end of the cylinder. Let us denote the end fragments by A and Z, and the fragments adjoining A and Z by B and Y respectively. The geometry of the fragmentation pattern indicates that the fragments A, B, Y and Z were emitted at angles of

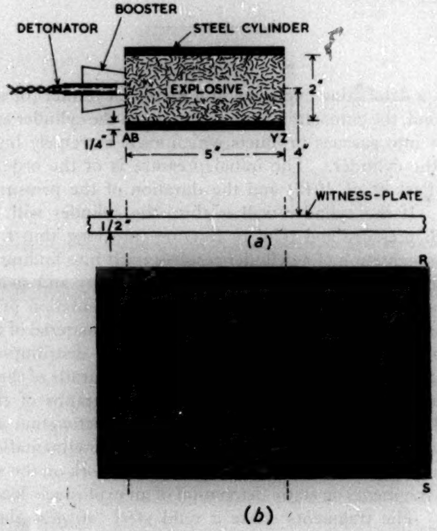


Figure 1 (*a*). Lay-out of the equipment for firing; (*b*) typical fragmentation pattern on a mild steel witness-plate. The dotted lines show the projected area of the steel cylinder normal to the witness-plate.

−2°, 3°, 8° and 45° to the normal to the inner surface of the cylinder respectively. The arc RS is obviously due to the impact of fragments arising as a result of the 'corner fracturing' plus subsequent expansion (Singh 1956).

2.2. Steel rings $\frac{1}{8}$ in. thick and 2 in. long were fitted snugly over a high explosive cartridge (2 in. diameter and 14 in. long). Thus the equipment consisted of confined (having steel ring) and bare charges alternately. The lay-out of the equipment for firing is shown in figure 2 (*a*). The fragmentation pattern obtained by initiating the charge at one end is shown in figure 2 (*b*). Each confined charge gives a fragmentation pattern on the witness-plate. The geometry of the pattern indicates that in a confined annular ring at the centre of the explosive charge, the fragments A, Y and Z are emitted at mean angles of 5°, 8° and 18° to the normal to the inner surface of the steel ring respectively. The fragments emitted at 18° obviously arise as a result of the 'corner fracturing' and their direction of emission is affected by the gases from the bare charge adjoining the fragment.

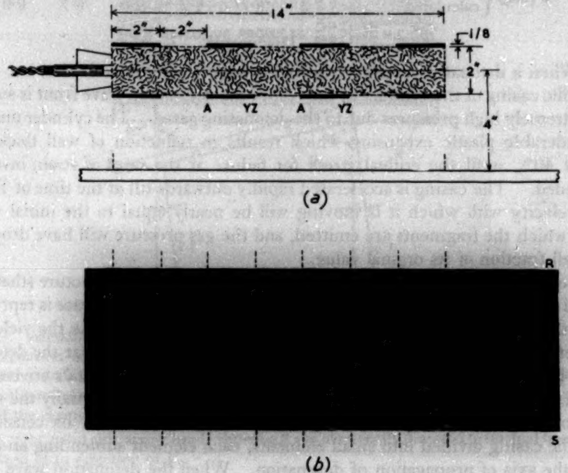


Figure 2. (*a*) Lay-out of the equipment for firing; (*b*) typical fragmentation pattern on a mild steel witness-plate. The dotted lines show the projected areas of the steel annular rings normal to the witness-plate.

2.3. Steel plates were arranged tangentially along circular arcs having a radius of 10 feet and the explosively loaded cylinder was detonated at the centre. As the distance of the witness-plate from the steel cylinder was considerably more than the length of the cylinder, all angles were measured from the centre of the cylinder. The number of 'troughs' on the steel plates in horizontal zones indicate that most of the fragments are emitted at angles of 6° to 9° to the normal.

2.4. The explosive was cast in copper pipes, with one face cut at an angle α where α is the angle between an element of casing and the axis of propagation of detonation. A thin steel plate was fixed to the cut face of the explosive. When the charge was fired the steel plate broke into fragments which struck the witness-plate and produced on it a fragmentation pattern. The angle δ between the direction of projection of a fragment and the normal to the inner surface of the metal casing was calculated from the geometry of the fragmentation pattern and is given in table 2.

Table 2. Experimental and Calculated δ

α (deg)		0	15	22·5	40	60	90
δ (deg)	experimental	7·5	9·0	9·4	9·6	6·6	1·5
	calculated	7·2	8·7	9·2	9·0	6·7	0·0

§ 3. Discussion

When a detonating explosive in a thin-walled cylinder is fired at one end the metallic casing of the cylinder in the rear of the detonation wave front is subjected to extremely high pressures due to the detonating gases. The cylinder undergoes considerable plastic expansion which results in reduction of wall thickness of about 40% until the critical stress for failure at the rates of strain involved is exceeded. The casing is accelerated rapidly outwards till at the time of fracture the velocity with which it is moving will be nearly equal to the initial velocity with which the fragments are emitted, and the gas pressure will have dropped to a small fraction of its orginal value.

Let V_0 represent the velocity of the casing at the time of fracture (that is, the initial velocity of the fragments). It is supposed that the metal case is represented by a thin flexible sheath of the same mass per unit length. As the yield stress of metallic casing is small compared with the stresses arising near the detonation wave front, the strength of the metal is neglected and these metals are treated as perfect fluids. The collapse mechanism assumed here is essentially the same as that of Birkhoff *et al.* (1948) for a wedge-shaped liner. Let us consider the metallic casing divided into small elements, each element subtending an angle α with the axis of propagation of detonation. When the detonation wave sweeps this element it is projected with a velocity V_0 and its direction of flight makes an angle δ with the normal to the element. V_0 and δ are given by the expressions (Singh 1955 b, Pugh *et al.* 1952)†

$$V_0 = \tfrac{1}{4} U_{\mathrm{d}}(1+\sin\alpha) \qquad \ldots\ldots(1)$$

$$\delta = \sin^{-1}(V_0 \cos\alpha / 2U_{\mathrm{d}}) \qquad \ldots\ldots(2)$$

where U_{d} is the velocity of detonation of the explosive. Combining equations (1) and (2), we have

$$\delta = \sin^{-1}(\tfrac{1}{8}\cos\alpha + \tfrac{1}{16}\sin 2\alpha). \qquad \ldots\ldots(3)$$

When α equals 90° (i.e. the plate is at right angles to the axis of propagation of detonation), the element is projected with a velocity $\frac{1}{2}U_{\mathrm{d}}$ and is normal to the original casing element. A cylinder is a limiting case of a wedge where the element is parallel to the axis of propagation of detonation and α equals 0°. Figure 3 shows the metallic casing at any instant having undetonated explosive ahead of the detonation wave and collapsed surface of the casing in the rear of the detonation wave. The element is projected with a velocity $\frac{1}{4}U_{\mathrm{d}}$ and its direction of projection makes an angle of $\sin^{-1}(V_0/2U_{\mathrm{d}})$ with the normal.

---

† V_0 depends upon wall thickness, ratio of weight of explosive to weight of metal and properties of the metal of the casing. As a first approximation, these parameters are neglected in equation (1).

The calculated δ by equation (3) is given in table 2. Similar values of δ were obtained by Kothari and Singh (1953) and Singh (1955 a) from the fragmentation patterns of 'inverted' conical liners having half angles α of $22\frac{1}{2}°$ and 40°. The quantitative agreement is good between experimental and calculated values of δ.

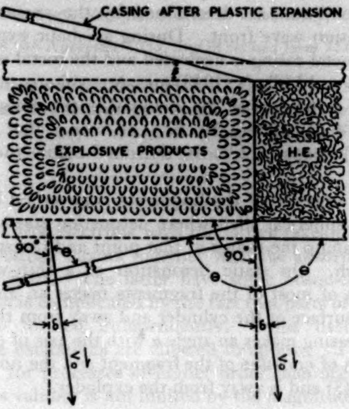


Figure 3. Metallic casing at any instant showing undetonated explosive and plastic expansion of casing in the rear of detonation wave front. The element P moves with velocity V_0 in a direction bisecting the angle between the original casing and the collapsed surface †, or making an angle δ with the normal to the original inner surface of the casing. $\delta = \frac{1}{2}\pi - \theta = \sin^{-1}(V_0/2U_d)$.

† On right-hand side arrow heads showing θ are incorrectly directed to normal to original wall instead of to direction of V_0.

Equation (3) also indicates that δ is independent of the thickness of casing and velocity of detonation of the explosive. The experiments described in § 2.1 and a very limited number of firings with different explosives (60 RDX, 40 TNT and 30 CE, 70 TNT) also support the above view.

The flash radiographs (Clark 1949) of 20 mm high explosive shell indicate that on detonation from one end the central portion of the casing becomes 'balloon-shaped' prior to rupture while each end retains approximately its original diameter. This bulging suggests that some of the casing at the initiator end is projected backwards (having a negative value of δ) towards the point of initiation. Figures 1 and 2 also show that the fragments arising from the casing at the initiator end have a negative value of δ.

The thickness of a fragment retaining parts of the original inner and the outer surfaces of a $\frac{1}{2}$ in. wall cylinder is about 0·32 in. Assuming steel to be incompressible and neglecting any longitudinal expansion, the conservation of mass of the original cylinder and that at the end of plastic expansion suggests that the inside diameter increases by as much as 80% before the case breaks into fragments. Davis and Parker (1948), in their experiments with thin-walled tubes tested at $-138°$F with internal pressure only, found that the tubes exhibited up to 50% plastic expansion before fractures occurred by cleavage. Mott also considered that for cylindrical steel shell cases considerable plastic expansion, as much as 50%, occurred before the case broke into fragments.

It can be safely assumed that the products of detonation expand adiabatically and that the initial conditions are determined by the pressure, temperature and volume in the detonation wave front. During adiabatic expansion of the gases, the strength of the metal casing is neglected and the metal is treated as a perfect fluid. Taking Jones and Miller's (1948) pressure–volume–temperature relations during the adiabatic expansion of the material in the detonation wave front of TNT at a loading density of $1{\cdot}5\,\mathrm{g\,cm^{-3}}$, it seems that the initial pressure ($15{\cdot}9 \times 10^{10}\,\mathrm{dyn\,cm^{-2}}$) drops to $2{\cdot}9 \times 10^{9}\,\mathrm{dyn\,cm^{-2}}$ before the case fractures into fragments.

The direction of emission of the fragments is governed largely by the shape of the shell. For a spherical shell where detonation starts at the centre, each fragment is emitted along the radius at that point and the spatial distribution of fragments is uniform. On static detonation of a thin-walled cylinder, the direction of emission of most of the fragments makes an angle of 7·5° with the normal to the inner surface of the cylinder and away from the exploder. If any element of the shell casing makes an angle α with the axis of detonation, the angle between the direction of emission of the fragment and the normal to that element is given by equation (3) and is away from the exploder.

ACKNOWLEDGMENTS

The author is grateful to Professor D. S. Kothari, Scientific Adviser to the Ministry of Defence, for his encouragement and interest throughout the course of this investigation. Thanks are also due to the Controller-General of Defence Production, Ministry of Defence, for facilities provided in connection with the above work.

REFERENCES

BIRKHOFF, G., MACDOUGALL, D. P., PUGH, E. M., and TAYLOR, G. I., 1948, *J. Appl. Phys.*, **19**, 563.
CLARK, J. C., 1949, *J. Appl. Phys.*, **20**, 363.
DAVIS, H. E., and PARKER, E. R., 1948, *J. Appl. Mech.*, **15**, 201.
JONES, H., and MILLER, A. R., 1948, *Proc. Roy. Soc.* A, **194**, 480.
KOTHARI, D. S., and SINGH, S., 1953, *Proc. Nat. Inst. Sci. India*, **19**, 507.
MOTT, N. F., 1947, *Proc. Roy. Soc.* A, **189**, 300.
PACK, D. C., EVANS, W. M., and JAMES, H. J., 1948, *Proc. Phys. Soc.*, **60**, 1.
PUGH, E. M., EICHELBERGER, R. J., and ROSTOKER, N., 1952, *J. Appl. Phys.*, **23**, 532.
SINGH, S., 1955 a, *Proc. Nat. Inst. Sci. India*, **21**, 129; 1955 b, *Proc. Phys. Soc.* B, **68**, 785; 1956, *Brit. J. Appl. Phys.*, **7**, **453**.

Effect of the Electron–Lattice Interaction on Phonon Frequencies

By P. G. HARPER
Division of Electrotechnology, National Standards Laboratory, Sydney, Australia

MS. received 9th *November* 1955, *and in revised form* 25*th May* 1956

Abstract. The interaction of free electrons with the lattice vibrations is treated, allowing only the virtual excitations of a single electron–hole pair to contribute to the phonon energy. Changes in sound velocity for the Fermi and Fröhlich states are compared.

§ 1. Introduction

The interaction of free electrons with the lattice takes place by absorption and emission of lattice quanta together with the absorption and emission of electron–hole pairs. The latter have some boson-like properties but their multiple excitation is complicated by the fact that pairs of different momenta cannot in general be formed independently. Some useful results follow, however, if only single excitations are allowed to appear. Though the method does not then account for processes of higher order than occur in second-order perturbation theory, its validity is not limited by the magnitude of the interaction constant. It is used here to compare changes in sound velocity for the Fermi and Fröhlich states.

§ 2. The Electron–Hole Description

The field theory notation being a suitable one, the notation of Fröhlich (1952) will be closely followed. The total Hamiltonian is written as

$$H=H_{\mathrm{el}}+H_L+H_{\mathrm{int}}, \qquad \ldots\ldots(1)$$

where $H_{\mathrm{el}}=\sum_{\mathbf{k}} a_{\mathbf{k}}{}^{\dagger}a_{\mathbf{k}}\epsilon_{\mathbf{k}}$ ($\epsilon_{\mathbf{k}}$ being the kinetic energy of a free electron), $H_L=\frac{1}{2}\sum_{\mathbf{q}}\hbar\omega_{\mathbf{q}}(b_{\mathbf{q}}b_{\mathbf{q}}{}^{\dagger}+b_{\mathbf{q}}{}^{\dagger}b_{\mathbf{q}})$ ($\omega_{\mathbf{q}}$ being the longitudinal phonon frequency) and $H_{\mathrm{int}}=i\sum_{\mathbf{k},\mathbf{q}}D_{\mathbf{q}}(b_{\mathbf{q}}a_{\mathbf{k}}{}^{\dagger}a_{\mathbf{k}-\mathbf{q}}-b_{\mathbf{q}}{}^{\dagger}a_{\mathbf{k}}{}^{\dagger}{}_{-\mathbf{q}}a_{\mathbf{k}})$, the first term expressing the absorption of a phonon of momentum $\mathbf{q}$ together with the translation of an electron from the state $\mathbf{k}-\mathbf{q}$ into $\mathbf{k}$. Regarding the latter as the emission of an electron–hole pair of energy $\hbar\omega=\epsilon_{\mathbf{k}}-\epsilon_{\mathbf{k}-\mathbf{q}}$ and momentum $\mathbf{q}$, it is convenient to introduce the operator $\rho_{\mathbf{q}}{}^{\dagger}{}_{,\omega}$ [‡] which produces such pairs. That is, $\rho^{\dagger}_{\mathbf{q},\omega}$ is a Fourier frequency component of $\rho_{\mathbf{q}}$ and consequently

$$\rho_{\mathbf{q}}{}^{\dagger}{}_{,\omega}=\sum_{\mathbf{k}} a_{\mathbf{k}}{}^{\dagger}a_{\mathbf{k}-\mathbf{q}}\delta(\epsilon_{\mathbf{k}}-\epsilon_{\mathbf{k}-\mathbf{q}}-\hbar\omega). \qquad \ldots\ldots(2)$$

[‡] This definition needs to be understood carefully. For arbitrary ω exact resonances in equation (3) will not in general occur, i.e. there will be no exact solutions of $\epsilon_{\mathbf{k}}-\epsilon_{\mathbf{k}-\mathbf{q}}=\hbar\omega$, where $k_x=(2\pi/L_x)t$, t integral, etc. The selection of exact Fourier components may be achieved in the following way. Suppose $\mathbf{q}$ (also discretely spaced) is given; one may then ask what are the permitted values of ω which allow true resonances ? Now the three orthogonal directions in which the cyclic boundary conditions are applied are arbitrary. Take one to be in the $\mathbf{q}$ direction. Since the condition $\epsilon_{\mathbf{k}}-\epsilon_{\mathbf{k}-\mathbf{q}}=\hbar\omega$ defines solutions for ω to lie on a plane perpendicular to $\mathbf{q}$ it is clearly possible to choose ω so that this plane coincides with a (1, 0, 0) plane of the lattice points. Every lattice point in this plane then provides an exact resonance. A neighbouring value of ω is obtained by translating the plane parallel to itself one $\mathbf{k}$-spacing so that the ω-spacing is given by

$$\Delta(\hbar\omega)=q(\hbar^2/2m)(2\pi/L).$$

Thus for fixed q, ω-values are uniformly spaced with a density $(L/2\pi)(m/\hbar q)$.

Similarly,

$$\rho_{\mathbf{q},\omega} = \sum_{\mathbf{k}} a_{\mathbf{k}-\mathbf{q}}{}^{\dagger} a_{\mathbf{k}} \delta(\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}-\mathbf{q}} - \hbar\omega). \qquad \ldots\ldots(3)$$

The effect of $\rho_{\mathbf{q}}{}^{\dagger}{}_{,\omega}$ on a state specified by occupation numbers $n_{\mathbf{k}}$[‡] is to produce a linear combination of such states, the number of pairs, $(\mathbf{q}, \omega)$, in each being increased by one; similarly, $\rho_{\mathbf{q},\omega}$ decreases by one the number of pairs.

In terms of $\rho_{\mathbf{q},\omega}$, $\rho_{\mathbf{q}}{}^{\dagger}{}_{,\omega}$, H_{int} becomes

$$i \sum_{\mathbf{q},\omega} D_{\mathbf{q}} (b_{\mathbf{q}} \rho_{\mathbf{q}}{}^{\dagger}{}_{,\omega} - b_{\mathbf{q}}{}^{\dagger} \rho_{\mathbf{q},\omega}). \qquad \ldots\ldots(4)$$

The following simple properties of ρ, $\rho^{\dagger}$ may be verified directly from the definitions (2) and (3).

(i) $$\rho_{\mathbf{q},\omega} = \rho^{\dagger}_{-\mathbf{q},-\omega}$$

(ii) $$[\rho_{\mathbf{q},\omega}, H_{\text{el}}] = \hbar\omega \rho_{\mathbf{q},\omega}; \quad [\rho_{\mathbf{q}}{}^{\dagger}{}_{,\omega}, H_{\text{el}}] = -\hbar\omega \rho_{\mathbf{q}}{}^{\dagger}{}_{,\omega}$$

(iii) $$[\rho_{\mathbf{q},\omega}, \rho_{\mathbf{q},\omega'}] = 0$$

(iv) $$[\rho_{\mathbf{q},\omega}, \rho^{\dagger}_{\mathbf{q},\omega'}] = \delta_{\omega,\omega'} \sum_{\mathbf{k}} (n_{\mathbf{k}-\mathbf{q}} - n_{\mathbf{k}}) \delta(\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}-\mathbf{q}} - \hbar\omega) \equiv \delta_{\omega,\omega'} \xi_{\mathbf{q},\omega}$$

(v) $$[\rho_{\mathbf{q},\omega}, \xi_{\mathbf{q},\omega'}] = -2\delta_{\omega,\omega'} \rho_{\mathbf{q},\omega}; \quad [\rho_{\mathbf{q},\omega}, \xi_{\mathbf{q},\omega'}] = 2\delta_{\omega,\omega'} \rho^{\dagger}_{\mathbf{q},\omega}$$

(vi) $$[\xi_{\mathbf{q},\omega}, H_{\text{int}}] = 0.$$

There are not many further useful relations between ρ and $\rho^{\dagger}$ of different $\mathbf{q}$ values. One may be mentioned,

$$[\rho_{\mathbf{q}+\mathbf{q}',\omega+\omega'}, [\rho_{\mathbf{q},\omega}, \rho_{\mathbf{q}',\omega'}]] = 0$$

which will not, however, be needed here.

Since $\xi_{\mathbf{q},\omega}$ is Hermitian, its diagonal representation may be used.[||] The corresponding eigenstates are generated by successive operations with $\rho_{\mathbf{q}}{}^{\dagger}{}_{,\omega}$ upon a single-electron state. Thus, from (v), $(\xi_n - \xi_{n'} + 2)\rho_{n,n'} = 0$ (for some $\mathbf{q}$, ω) so that $\rho_{n,n-1}$ may be taken as the only non-vanishing matrix element of ρ. Consequently,

$$\xi_n = M - 2n \qquad \ldots\ldots(5)$$

where n is some integer. It is shown in the Appendix that M (a function of q, ω), is proportional to L^2. It also clearly depends on the single-electron state chosen. The actual values of $\rho_{n,n-1}$, and $\rho_n{}^{\dagger}{}_{,n+1}$ are obtained, either algebraically from (iv), or otherwise, as

$$\left.\begin{aligned} \rho_{n,n-1} &= \{n(M-n+1)\}^{1/2} \\ \rho_n{}^{\dagger}{}_{,n+1} &= \{(n+1)(M-n)\}^{1/2} \end{aligned}\right\} \qquad \ldots\ldots(6)$$

Returning to the problem of interaction, a simple illustration of the use of pairs, which also serves as a guide to the more difficult problem, is that of a static sinusoidal potential imposed on the free electrons. The Hamiltonian is then

$$H = H_{\text{el}} + \sum_{\omega} (a\rho_{\omega}{}^{\dagger} + a^{*}\rho_{\omega})$$

where the momentum index $\mathbf{q}$ (suppressed for convenience) is also the wave vector of the potential, and a is its amplitude. Then taking time derivatives of ρ_{ω}, $\rho_{\omega}{}^{\dagger}$

$$[\rho, H] = \hbar\omega\rho + a\xi$$

$$-[\rho^{\dagger}, H] = \hbar\omega\rho^{\dagger} + a^{*}\xi.$$

[‡] For convenience referred to as a single-electron state.

[||] It is easily verified that $\frac{1}{2}\Sigma_{\mathbf{q},\omega}\, \hbar\omega\, \xi_{\mathbf{q},\omega} = -H_{\text{el}}\Sigma_{\mathbf{k}} + N\Sigma_{\mathbf{k}}\epsilon_{\mathbf{k}}$, where both sums extend over all $\mathbf{k}$-values and are therefore infinite.

Regarding ξ as an independent variable,

$$[\xi, H] = -2a\rho^\dagger + 2a^*\rho.$$

If E_n are the eigenvalues of H,

$$\left.\begin{aligned}(\hbar\omega - E_{n,n'})\rho_{n,n'} + a\xi_{n,n'} &= 0\\ (\hbar\omega + E_{n,n'})\rho_{n,n'}{}^\dagger + a^*\xi_{n,n'} &= 0\\ -2a\rho_{n,n'}{}^\dagger + 2a^*\rho_{n,n'} - E_{n,n'}\xi_{n,n'} &= 0\end{aligned}\right\} \quad \ldots\ldots(7)$$

where $E_{n,n'} = E_n - E_{n'}$. The equations (7) lead to the secular equation for $E_{n,n'}$

$$E_{n,n'}(E_{n,n'}{}^2 - 4|a|^2 - (\hbar\omega)^2) = 0. \quad \ldots\ldots(8)$$

The root $E_{n,n'} = 0$ shows that ρ, $\rho^\dagger$, ξ have finite diagonal elements. The energy spectrum consists of uniformly spaced levels (pair energies) of separation ΔE, where $(\Delta E)^2 = (\hbar\omega)^2 + 4|a|^2$. This expresses the energy gap introduced by the static wave into the free electron energy, $\omega = 0$ for example, (corresponding to $\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}-\mathbf{q}} = 0$) giving $\Delta E = \pm 2|a|$.

It is evident that a more complicated potential would lead to difficulties since the commutators $[\rho_{\mathbf{q},\omega}, \rho_{\mathbf{q}',\omega'}]$ would be introduced which have no simple properties. The same happens in the phonon interaction and the limitations of the method become apparent. Consider, however, the simpler problem in which only phonons of $\pm\mathbf{q}$ can be excited; the sinusoidal potential is now replaced by a travelling wave. An exact solution to this problem could be valuable even if only for its mathematical interest. Denote the phonon amplitudes by b_+ and b_-. Then proceeding as before and using (4),

$$\left.\begin{aligned}[\rho_\omega, H] &= \hbar\omega\rho_\omega + iD(b_+ - b_-{}^\dagger)\xi_\omega\\ [\rho_\omega{}^\dagger, H] &= -\hbar\omega\rho_\omega{}^\dagger - iD(b_- - b_+{}^\dagger)\xi_\omega\\ [b_+, H] &= \hbar s'qb_+ - iD\sum_\omega \rho_\omega\\ [b_+{}^\dagger, H] &= -\hbar s'qb_+{}^\dagger - iD\sum_\omega \rho_\omega{}^\dagger,\end{aligned}\right\} \quad \ldots\ldots(9)$$

and two more equations for b_-, $b_-{}^\dagger$,

$$[\xi_\omega, H] = 2iD\{(b_- - b_+{}^\dagger)\rho_\omega - (b_+ - b_-{}^\dagger)\rho_\omega{}^\dagger\}.$$

The expressions for $\partial\rho_\omega/\partial t$, $\partial\rho_\omega{}^\dagger/\partial t$ and $\partial\xi_\omega/\partial t$ are clearly non-linear, involving as they do the products $b\mathscr{E}$, $b\rho$ etc. In the diagonal representation (9) may be presented as

$$\left.\begin{aligned}&\rho_{n,n'} + iD\left\{(b_+ - b_-{}^\dagger)\sum_\omega \frac{\xi_\omega}{\hbar\omega - E_{n,n'}}\right\}_{n,n'} = 0, \text{ and its complex conjugate,}\\ &\qquad (b_+ - b_-{}^\dagger)_{n,n'} = \frac{2iD\hbar s'q}{(\hbar s'q)^2 - E_{n,n'}{}^2}\rho_{n,n'}, \text{ and its complex conjugate,}\\ &\qquad (\xi_\omega)_{n,n'}E_{n,n'} = 2iD\{(b_- - b_+{}^\dagger)\rho_\omega - (b_+ - b_-{}^\dagger)\rho_\omega{}^\dagger\}_{n,n'}\end{aligned}\right\} \ldots(10)$$

where $\rho = \sum_\omega \rho_\omega$. Putting $(b_+ - b_-{}^\dagger)_{n,n'} = B_{n,n'}$, the matrix B satisfies the equation,

$$\{(\hbar s'q)^2 - E_{n,n'}{}^2\}B_{n,n'} = 2\hbar s'qD_q{}^2\left\{B\sum_\omega \frac{\xi_\omega}{\hbar\omega - E_{n,n'}}\right\}_{n,n'}. \quad \ldots\ldots(11)$$

The right hand side of (11) will contain both diagonal and non-diagonal elements of ξ_ω. If their values were *assigned*, (11) would be a linear equation for B. However, the non-diagonal values are determined from the last equation of (10) by the non-linear terms $b\rho_\omega$ etc. No way of overcoming this difficulty has been found and it is concluded that not even for this apparently simple problem is a rigorous solution possible by such means.

It is evident that $\xi_{n,n'}$ $(n'\neq n)\propto D_q^2$; consider the approximation $\xi_{n,n'}=0$. Equation (11) becomes

$$(\hbar s'q)^2-(\hbar\Omega_{\mathbf{q}})^2=2\hbar s'qD_q^2\sum_\omega\frac{\xi_\omega}{\hbar\omega-\hbar\Omega_{\mathbf{q}}} \qquad \ldots\ldots(12)$$

where $B_{n,n'}=0$ except between adjacent levels separated by $E_n-E_{n'}=\hbar\Omega_{\mathbf{q}}$. This equation for the new phonon frequency $\Omega_{\mathbf{q}}$ though inexact is of more practical value than (11) since it can be taken to hold for independent excitations of all phonons $\mathbf{q}$, the interference terms $\rho_{\mathbf{q},\omega}$, $\rho_{\mathbf{q}',\omega'}$ having non-diagonal matrix elements also $\propto D_q^2$. In fact, as developed here, the retention of phonons at all can only be at the expense of abandoning higher order terms which ruin the independence of different $\mathbf{q}$ modes.

Consider a similar approximation applied to the static case. From the last equation of (7), $\xi_{n,n'}\propto a$, to lowest order. The second-order correction to $\hbar\omega$ is $E_{n,n'}=\hbar\omega+2|a|^2/\hbar\omega$, i.e., the usual perturbation correction. It is the exact consideration of the non-diagonal part of ξ which removes the resonance denominator and replaces the energy divergence by a finite discontinuity. The same is likely to be true of the dynamical case so that resonance transitions (real scatterings) occur only when $\xi_{n,n'}$ $(n'\neq n)$ can be disregarded. Now $\xi_{n,n'}$ is finite only between linearly combined single-electron states so that neglect of such matrix elements is related to the absence of correlation between single-electron momentum states. That (12) refers to one pair excited is obvious from the appearance of only a single energy $\hbar\omega$. It is also clear that higher order processes are contained in the abandoned $D_{\mathbf{q}}^2$ terms.

Diagonal values of ξ are undetermined by the last equation of (10) and indeed are independent of D. Since ξ is Hermitian, they are real and may be taken to be identical with the eigenvalues of ξ in the ξ-representation already mentioned. In what follows the ground state value $M(\mathbf{q},\omega)$ will be employed. Actually, regarding $\xi=[\rho,\rho^\dagger]$ as constant amounts to treating the pairs as boson particles, the matrix elements of ρ and $\rho^\dagger$ being nearly $M^{1/2}n^{1/2}$, $M^{1/2}(n+1)^{1/2}$ respectively, neglecting n against M which is proportional to L^2.

The summation over virtual pair energies of (12) is replaced by an integration, the density of ω-values being $(L/2\pi)(m/\hbar q)$ as seen. The denominator on the right-hand side of (12) vanishes at $\hbar\omega=\hbar\Omega_{\mathbf{q}}$ and this corresponds to real transitions. Note that the frequency of the scattered phonons is the true one, $\Omega_{\mathbf{q}}$, thus supporting the usual practice of using the measured sound velocity in the scattering matrix elements of Bloch theory. Such processes are not excluded by this treatment but as they are not under discussion, the integral will be replaced by its principal value.

Putting $\Omega_{\mathbf{q}}=sq$, equation (12) determines s as a function of s', C. If only the Fermi state were considered, the relation would be of no practical interest since s' is unobservable. It may be used, however, to compare sound velocities appropriate to (i) the Fermi state, and (ii) the shell state proposed by Fröhlich (1950) as the superconducting ground state. Perhaps one should not take the results too seriously since the assumption of independent phonons for the superconducting state is not established by these arguments.

§ 3. New Phonon Frequencies

3.1. *Fermi State*

Since $M(\mathbf{q}, \omega)$ is an odd function of ω, if the limits of ω are $\pm\omega_{\max}$,

$$\int_{-\omega(\max)}^{\omega(\max)} \frac{M d\omega}{\omega - s'q} = 2\int_{0}^{\omega(\max)} \frac{\omega M d\omega}{\omega^2 - (s'q)^2}.$$

Using the Fermi value of M given in the Appendix, and simplifying,

$$1 - \left(\frac{s}{s'}\right)^2 = 2\nu F' \left\{ \int_0^{1-\beta} \frac{u^2 du}{u^2 - (s/v)^2} + \frac{1}{4\beta}\int_{1-\beta}^{1+\beta} \frac{u\{1-(u-\beta)^2\} du}{u^2 - (s/v)^2} \right\},$$

where F' is the interaction parameter defined by $F' = \frac{3}{8}C^2/M_{\text{ion}}s'^2\zeta$ and $\beta = q/2k_0$ For $q \ll k_0$ the equation reduces to

$$1 - \left(\frac{s}{s'}\right)^2 = 2\nu F' \left\{ 1 + \frac{s}{2v}\ln\left(\frac{v}{s} - 1\right) \right\}.$$

Now s is the measured value so that $s \ll v$; neglecting s/v and using $F = F'(s'/s)^2$,

$$1 = (s/s')^2(1 + 2F)$$

which relates the observed s with the fictitious s'. Apart from the rather trivial difference that the condition $2\nu F' < 1$ ensures that s is now real rather than positive, the present results are equivalent to Fröhlich's (1952).

3.2. *Fröhlich Shell State*

This state may be supposed generated by successive operations on the Fermi state with $\rho^{\dagger}_{q,\omega_0}$ up to $n = M(\mathbf{q}, \omega_0) - 1$ for some fixed (arbitrary) value of q, and all orientations in a given (arbitrary) plane. This results in an isotropic gap $\hbar\omega_0$ separating the filled sphere from a concentric shell of energy depth $\hbar\omega_0$.

First consider $vq < \omega_0$; excitations across the gap are then impossible and therefore since $\hbar\omega_0$ is much less than the Fermi energy, M-values are the same as in the Fermi case. The Fröhlich value of M then does not differ at all from the Fermi value.

If now $vq > \omega_0$ (but $q \ll k_0$), the M-values, given in the appendix, are complicated. To the approximation used here, however, inspection of the integrals concerned shows that in lowest order, their value differs from the Fermi value by $-(\hbar\omega_0/\zeta)^2$ where $\zeta =$ Fermi energy. With a common value of s' for the two states, therefore (12) shows that the Fröhlich sound velocity for $q \ll k_0$ exceeds the Fermi sound velocity by the order of $(\hbar\omega_0/\zeta)^2$.

§ 4. Conclusions

The practical results of this paper relate to the change in sound velocity due to the electron–lattice interaction. The difficulties encountered, however, are not without interest. The simple case of electrons interacting with a sinusoidal lattice oscillation already emphasizes that the operator $\xi_{\mathbf{q},\omega}$ (which is a sum of Fermi factors, $n_{\mathbf{k}-\mathbf{q}}(1 - n_{\mathbf{k}})$) is certainly non-diagonal for true eigenstates and that the non-diagonal elements ignored here and in second-order perturbation treatments may remove the pole normally due to the vanishing of the resonance denominator $\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}-\mathbf{q}} + \hbar\omega_{\mathbf{q}}$. In that case, real transitions would be excluded. It seems, therefore, that further progress might be aided by starting not with a free electron gas, but correlated electron–lattice states, self-consistent values of ξ being derived. This would be in keeping with the view (Fröhlich 1954) that low-stage processes, such as given here, cannot deal effectively with the problem and that the 'ordering' should be introduced from the beginning.

ACKNOWLEDGMENT

The author is indebted to the referee for useful criticism.

APPENDIX

The equation $\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}-\mathbf{q}} = \hbar\omega$ determines $\mathbf{k}$, $\mathbf{k}-\mathbf{q}$ to lie on planes normal to $\mathbf{q}$ and at distances $(m\omega/\hbar q) \pm \frac{1}{2}q$ from the origin $k=0$. Consider first the Fermi ground state.

Then

$$\sum_{\mathbf{k}} n_{\mathbf{k}-\mathbf{q}} \delta(\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}-\mathbf{q}} - \hbar\omega) = \left(\frac{L}{2\pi}\right)^2 \pi \left\{ k_0{}^2 - \left(\frac{m\omega}{\hbar q} - \tfrac{1}{2}q\right)^2 \right\}$$

(where the right-hand side is $(L/2\pi)^2$ times the area of the plane intercepted by the Fermi sphere of radius k_0) provided $\{(m\omega/\hbar q) - \frac{1}{2}q\} < k_0$; the expression is equal to zero otherwise. Similarly

$$\sum_{\mathbf{k}} n_{\mathbf{k}} \delta(\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}-\mathbf{q}} - \hbar\omega) = \left(\frac{L}{2\pi}\right)^2 \pi \left\{ k_0{}^2 - \left(\frac{m\omega}{\hbar q} + \tfrac{1}{2}q\right)^2 \right\}.$$

Therefore $M(q, \omega)$, clearly an odd function of ω, is given by

$$M(q,\omega) = \begin{cases} \pi\left(\dfrac{L}{2\pi}\right)^2 \dfrac{2m\omega}{\hbar}, & 0 < \omega < vq\left(1 - \dfrac{q}{2k_0}\right) \\ \pi\left(\dfrac{L}{2\pi}\right)^2 \left\{ k_0{}^2 - \left(\dfrac{m\omega}{\hbar q} - \tfrac{1}{2}q\right)^2 \right\}, & vq\left(1 - \dfrac{q}{2k_0}\right) < \omega < vq\left(1 + \dfrac{q}{2k_0}\right) \\ 0, & \text{greater values.} \end{cases}$$

where $v = \hbar k_0/m$, the Fermi velocity.

The Fröhlich state $M(q, \omega)$ may be similarly considered. Only the case $vq > 2\omega_0$ will be written down here.

$$\begin{aligned} M(q,\omega) = \pi\left(\frac{L}{2\pi}\right)^2 \times \frac{2m\omega}{\hbar}, &\qquad 0 < \omega < v_{-}q(1 - q/2k) \\ k^2 - \left(\frac{m\omega}{\hbar q} - \tfrac{1}{2}q\right)^2, &\qquad v_{-}q(1 - q/2k_{-}) < \omega < vq(1 - q/2k_0) \\ \frac{2m}{\hbar}(\omega - \omega_0), &\qquad vq(1 - q/2k_0) < \omega < v_{+}q(1 - q/2k_{+}) \\ k_0{}^2 - \left(\frac{m\omega}{\hbar q} - \tfrac{1}{2}q\right)^2, &\qquad v_{+}q(1 - q/2k_{+}) < \omega < v_{-}q(1 + q/2k_{-}) \\ \frac{2m\omega_0}{\hbar}, &\qquad v_{-}q(1 - q/2k) < \omega < vq(1 + q/2k_0) \\ k_{+}{}^2 - \left(\frac{m\omega}{\hbar q} - \tfrac{1}{2}q\right)^2, &\qquad vq(1 + q/2k_0) < \omega < v_{+}q(1 + q/2k_{+}) \\ 0, &\qquad \text{greater values of } \omega \end{aligned}$$

where $v_{+} = \hbar k_{\pm}/m$, $k_{\pm}{}^2 = k_0{}^2 \pm 2m\omega_0/\hbar$.

REFERENCES

FRÖHLICH, H., 1950, *Phys. Rev.*, **79**, 845; 1952, *Proc. Roy. Soc.*, **215**, 291; 1954, *Proceedings of the International Conference of Theoretical Physics*, Kyoto and Tokyo, September 1953 (Tokyo: Science Council of Japan).

Anisotropic Galvanomagnetic Effects in Semiconductors

By J. R. DRABBLE and R. WOLFE

Research Laboratories of The General Electric Company Limited, Wembley, Middx.

Communicated by the G.E.C.; MS. received 11*th May* 1956

Abstract. The 'many-valley' theory of the galvanomagnetic effects in cubic semiconductors proposed by Abeles and Meiboom is extended to include semiconductors belonging to any crystal class. The theory is applied to the crystal class $\bar{3}$ m. Explicit expressions for the components of the conductivity, Hall conductivity and magnetoconductivity tensors are derived for a six-valley model in which the energy extrema lie on the reflection planes in **k**-space. These expressions are consistent with preliminary measurements on a single crystal of n-type bismuth telluride in which all the galvanomagnetic effects are anisotropic.

---

§ 1. INTRODUCTION

Recent measurements in these laboratories of the electrical properties of bismuth telluride (to be published) have shown that the conductivity and Hall coefficient are markedly anisotropic in both n-type and p-type single crystals. The conductivity measured in the cleavage planes is approximately four times greater than that measured perpendicular to these planes and the Hall coefficient varies with crystal orientation by a factor of two or three. The ratios, which differ in the two types of material, remain constant over the temperature range 100°K to 200°K. The Hall coefficients remain practically constant in this range, indicating that only one type of carrier is present.

In a simple model of a semiconductor, the surfaces of constant energy in **k**-space are assumed to be ellipsoids centred at the origin, and the relaxation time is assumed to be a function of energy only. Using such a model, and assuming that there are two types of charge carrier, Jones (1936) was able to explain the anisotropic Hall coefficient and conductivity of bismuth. However, with only one type of carrier present, the above model leads to a Hall coefficient which is independent of crystal orientation although the conductivity is anisotropic (Bronstein 1932, Johnson 1949).

A theory of the galvanomagnetic effects based on the assumptions that there are several energy extrema at equivalent points in **k**-space and that the surfaces of constant energy are ellipsoids centred at these points, has been developed by Abeles and Meiboom (1954) and Herring (1955). They applied this 'many-valley' model to cubic semiconductors. In such materials the conductivity and Hall coefficient are isotropic. However, for materials with non-cubic symmetry, this theory does predict that these properties may be anisotropic.

In this paper, the 'many-valley' model is applied to the crystal class $\bar{3}$ m to which bismuth telluride belongs (Lange 1939). The theory involves several parameters which are related to the shape and orientation of the energy ellipsoids as well as the density of carriers and the relaxation time. Measurements of the resistivity and Hall coefficient tensor components are not sufficient to allow a determination of all these parameters. The theory is therefore extended to include the low field magnetoresistance and related effects. Of the various arrangements of valleys which are consistent with the symmetry of this crystal class, the one in which there are six valleys with centres on the three reflection planes is found to give results which agree qualitatively with preliminary measurements on n-type bismuth telluride. Therefore, explicit expressions for the conductivity, Hall conductivity and magnetoconductivity tensor components are given for this case. The methods described can be applied to semiconductors or metals belonging to any crystal class and with any consistent arrangement of valleys.

§ 2. PHENOMENOLOGICAL THEORY

The general expression for the current density **J** in terms of the electric field **E** and the magnetic field **H** is

$$J_i = \sigma_{ij}E_j + \sigma_{ijk}E_jH_k + \sigma_{ijkl}E_jH_kH_l + \dots \qquad \dots\dots(1)$$

where the usual summation convention applies. (Higher order terms are negligible in most experiments.) This equation defines the conductivity, Hall conductivity and magnetoconductivity tensors. The resistivity, Hall coefficient and magnetoresistivity tensors are defined by the inverse relation

$$E_i = \rho_{ij}J_j - \rho_{ijk}J_jH_k + \rho_{ijkl}J_jH_kH_l + \dots \qquad \dots\dots(2)$$

The number of independent components of these tensors and the relations among all of the non-zero components are determined by the symmetry of the crystal and the Onsager reciprocal relations. Juretschke (1955) has applied this phenomenological theory to crystals which belong to the class 3 m and Fumi (1952) has shown that the same results are obtained for the class $\bar{3}$ m. Referred to an orthogonal set of axes in which Oz is parallel to the three-fold axis of the crystal and the (x, z) plane is one of the three reflection planes, there are twelve independent components of the resistivity tensors ρ_{ij}, ρ_{ijk} and ρ_{ijkl}. These components and their relation to the independent components of the conductivity tensors σ_{ij}, σ_{ijk} and σ_{ijkl}, are

$$\left.\begin{array}{ll}
\rho_{11} = 1/\sigma_{11} & \rho_{1122} = -\sigma_{1122}/\sigma_{11}^2 - \sigma_{231}^2/\sigma_{11}^2\sigma_{33} \\
\rho_{33} = 1/\sigma_{33} & \rho_{1133} = -\sigma_{1133}/\sigma_{11}^2 - \sigma_{123}^2/\sigma_{11}^3 \\
\rho_{123} = \sigma_{123}/\sigma_{11}^2 & \rho_{3311} = -\sigma_{3311}/\sigma_{33}^2 - \sigma_{231}^2/\sigma_{11}\sigma_{33}^2 \\
\rho_{231} = \sigma_{231}/\sigma_{11}\sigma_{33} & \rho_{1123} = -\sigma_{1123}/\sigma_{11}^2 \\
\rho_{1111} = -\sigma_{1111}/\sigma_{11}^2 & \rho_{2311} = -\sigma_{2311}/\sigma_{11}\sigma_{33} \\
\rho_{3333} = -\sigma_{3333}/\sigma_{33}^2 & \rho_{2323} = -\sigma_{2323}/\sigma_{11}\sigma_{33} + \frac{1}{2}\dfrac{\sigma_{123}\sigma_{231}}{\sigma_{11}^2\sigma_{33}}
\end{array}\right\} \dots\dots(3)$$

These equations remain valid when ρ and σ are interchanged throughout. The other non-vanishing components are all related to these; in particular

$$\rho_{22} = \rho_{11} \quad \text{and} \quad \rho_{312} = \rho_{231}.$$

Also

$$\left.\begin{array}{l}
\rho_{ij} = 0 \quad \text{if} \quad i \neq j \\
\rho_{ijk} = -\rho_{jik} \\
\rho_{ijkl} = \rho_{jikl} = \rho_{ijlk} = \rho_{jilk}.
\end{array}\right\} \dots\dots(4)$$

In the following sections it is shown how to obtain expressions for the independent components of the conductivity tensors on the basis of a many-valley model.

In the usual experimental procedure the direction of the current flow is fixed and the resistivity components are measured directly. The equations (3) are therefore used in comparing the theory with experiment. Most of the tensor components are invariant under coordinate transformations which leave the direction of the three-fold axis unchanged. Therefore, in measurements of resistivity, Hall coefficient and longitudinal and transverse magnetoresistance effects, all directions in the cleavage planes are equivalent.

§ 3. ASSUMPTIONS

The model to be discussed is based on the following assumptions:

(*a*) Only one type of current carrier is present. These carriers are assumed to be electrons throughout this work but the translation to holes and the extension to a two-carrier model are elementary.

(*b*) The minimum energy occurs at a number of equivalent points $\mathbf{k}_0^{(i)}$ in $\mathbf{k}$-space. These points are related by the symmetry operations of the crystal.

(*c*) Surfaces of constant energy in $\mathbf{k}$-space are sets of similar ellipsoids centred about these points. With a suitable choice of axes, to be referred to as the principal axes of the *i*th valley, the energy E in the vicinity of $\mathbf{k}_0^{(i)}$ may be written

$$E=E_0+\frac{\hbar^2}{2}\left[\frac{(k_1^{(i)}-k_{10}^{(i)})^2}{m_1}+\frac{(k_2^{(i)}-k_{20}^{(i)})^2}{m_2}+\frac{(k_3^{(i)}-k_{30}^{(i)})^2}{m_3}\right] \quad \ldots\ldots(5)$$

where E_0 is the minimum energy in the band and m_1, m_2 and m_3 are the principal effective masses. (m_1^{-1}, m_2^{-1} and m_3^{-1} are the principal values of the reciprocal mass tensor.) This is the 'simple many-valley' model of Herring.

(*d*) A relaxation time τ exists and may be written in the form:

$$\tau=bE^{-\lambda} \quad \ldots\ldots(6)$$

where the index λ and the quantity b depend upon the scattering mechanism. Generally b will be a function of temperature. The validity of this equation has been discussed by Herring (1955).

§ 4. CONTRIBUTIONS TO THE CONDUCTIVITY TENSORS FROM A SINGLE VALLEY

For a single valley (the *i*th), the contributions to the conductivity tensors of equation (1) are given by the following expressions:

$$\sigma_{pq}^{(i)}=\frac{-e^2}{4\pi^3\hbar^2}\int\tau\frac{\partial f_0}{\partial E}\frac{\partial E}{\partial k_p}\frac{\partial E}{\partial k_q}\,d\mathbf{k} \quad \ldots\ldots(7a)$$

$$\sigma_{pqr}^{(i)}=\frac{-e^3}{4\pi^3\hbar^4c}\int\tau\frac{\partial f_0}{\partial E}\frac{\partial E}{\partial k_p}\frac{\partial E}{\partial k_l}\frac{\partial}{\partial k_m}\left(\tau\frac{\partial E}{\partial k_q}\right)\epsilon_{rlm}\,d\mathbf{k} \quad \ldots\ldots(7b)$$

$$\sigma_{pqrs}^{(i)}=\frac{-e^4}{4\pi^3\hbar^6c^2}\int\tau\frac{\partial f_0}{\partial E}\frac{\partial E}{\partial k_p}\frac{\partial E}{\partial k_j}\frac{\partial}{\partial k_l}\left[\tau\frac{\partial E}{\partial k_m}\frac{\partial}{\partial k_n}\left(\tau\frac{\partial E}{\partial k_q}\right)\right] \times\tfrac{1}{2}[\epsilon_{ljs}\epsilon_{nmr}+\epsilon_{ljr}\epsilon_{nms}]\,d\mathbf{k}. \quad \ldots\ldots(7c)$$

In these expressions, $-e$ is the charge of the electron, c is the velocity of light, $\hbar$ is Planck's constant divided by 2π, f_0 is the equilibrium distribution function, ϵ_{pqr} is the permutation symbol and E is the energy given by equation (5).

These expressions have been given by Abeles and Meiboom (1954) except that they used an unsymmetrical form for the magnetoconductivity tensor in which $\sigma_{pqrs} \neq \sigma_{pqsr}$. However, as pointed out by Herring (1955) only the symmetrical part is of physical significance.

With the assumptions of § 3, the evaluation of the above integrals leads to the expressions:

$$\sigma_{pq}^{(i)} = \frac{A}{m_p} \delta_{pq} \qquad \text{......(8 a)}$$

$$\sigma_{pqr}^{(i)} = \frac{-B}{m_p m_q} \epsilon_{pqr} \qquad \text{......(8 b)}$$

$$\sigma_{pqrs}^{(i)} = \frac{C}{m_p m_q m_t} [\epsilon_{tps}\epsilon_{qtr} + \epsilon_{tpr}\epsilon_{qts}] \qquad \text{......(8 c)}$$

where the superscript (i) is used to denote tensor components relative to the principal axes of the ith valley, and A, B and C are parameters whose values depend upon the form of f_0 and also upon b and λ in the relation $\tau = bE^{-\lambda}$.

For a non-degenerate semiconductor, it is found that

$$A = \frac{4}{3\pi^{1/2}} n^{(i)} e^2 b (kT)^{-\lambda} \Gamma\left(\frac{5}{2} - \lambda\right) \qquad \text{......(9 a)}$$

$$B = \frac{4}{3\pi^{1/2}} n^{(i)} \frac{e^3}{c} b^2 (kT)^{-2\lambda} \Gamma\left(\frac{5}{2} - 2\lambda\right) \qquad \text{......(9 b)}$$

$$C = \frac{2}{3\pi^{1/2}} n^{(i)} \frac{e^4}{c} b^3 (kT)^{-3\lambda} \Gamma\left(\frac{5}{2} - 3\lambda\right) \qquad \text{......(9 c)}$$

where k is Boltzmann's constant, T is the absolute temperature, $n^{(i)}$ is the density of electrons occupying states in the ith valley and $\Gamma(x)$ is the gamma function.

For a highly degenerate semiconductor, it is found that

$$A_{\mathrm{D}} = \frac{e^2}{3\pi^2} b \left(\frac{2m^*}{\hbar^2}\right)^{3/2} (\zeta)^{3/2-\lambda} \qquad \text{......(10 a)}$$

$$B_{\mathrm{D}} = \frac{e^3}{3\pi^2} \frac{b^2}{c} \left(\frac{2m^*}{\hbar^2}\right)^{3/2} (\zeta)^{3/2-2\lambda} \qquad \text{......(10 b)}$$

$$C_{\mathrm{D}} = \frac{e^4}{6\pi^2} \frac{b^3}{c^2} \left(\frac{2m^*}{\hbar^2}\right)^{3/2} (\zeta)^{3/2-3\lambda} \qquad \text{......(10 c)}$$

where ζ is the position of the Fermi level measured from the band edge and $m^* = (m_1 m_2 m_3)^{1/3}$.

For any value of λ these three quantities are related by the equation

$$B_{\mathrm{D}}^2 / A_{\mathrm{D}} C_{\mathrm{D}} = 2. \qquad \text{......(11)}$$

This relation is found to be useful in the comparison of theory with experiment. It is also useful to note that:

$\sigma_{pq}^{(i)} = 0$ unless $p = q$
$\sigma_{pqr}^{(i)} = 0$ unless p, q, r are all different
$\sigma_{pqrs}^{(i)} = 0$ unless p, q, r, s are equal in pairs.

§ 5. TRANSFORMATION AND SUMMATION OF TENSOR COMPONENTS

The conductivity tensors for the crystal as a whole are obtained by transforming the contributions from the separate valleys to a common coordinate system and summing over all the valleys. These valleys are related by the symmetry operations of the crystal. Therefore, if the orientation of one valley is specified, the orientation of all the others can be determined.

The symmetry operations of the crystal can be represented by a group of three-dimensional real orthogonal matrices R^{ij}. These may be considered as the coordinate transformations which carry the valley (i) into each of the other valleys (j). Let the matrices of direction cosines which specify the transformations of coordinates from the ith and jth valleys to the principal crystal axes (defined in § 2) be (a^i_{pq}) and (a^j_{pq}) respectively. Then the relation between these matrices is given by

$$a^j_{pq} = R^{ij}_{ps} a^i_{sq}. \qquad \text{......(12)}$$

The tensors associated with the jth valley are referred to the crystal axes by the usual tensor transformations; e.g.

$$\sigma^{(jc)}_{pq} = a^i_{pt} a^j_{qu} \sigma^{(j)}_{tu}. \qquad \text{......(13)}$$

The total conductivity tensor is obtained by summation over all the valleys

$$\sigma_{pq} = \sum_j \sigma^{(jc)}_{pq}. \qquad \text{......(14)}$$

Equations (12), (13) and (14) give

$$\sigma_{pq} = \sum_j R^{ij}_{pk} a^i_{kt} R^{ij}_{ql} a^i_{lu} \sigma_0{}^{(j)}_{tu} \qquad \text{.....(15)}$$

where the summation convention applies to the subscripts. Since the valleys are all similar, $\sigma^{(j)}_{tu} = \sigma^{(i)}_{tu}$, and the summation over valleys involves only the terms R^{ij}_{pk} and R^{ij}_{ql}. Therefore

$$\sigma_{pq} = \left\{ \sum_j R^{ij}_{kp} R^{ij}_{ql} \right\} a^i_{kt} a^i_{lu} \sigma^{(i)}_{tu}. \qquad \text{.....(16)}$$

Similarly, the total magnetoconductivity tensor is

$$\sigma_{pqrs} = \left\{ \sum_j R^{ij}_{pk} R^{ij}_{ql} R^{ij}_{rm} R^{ij}_{sn} \right\} a^i_{kt} a^i_{lu} a^i_{mv} a^i_{nw} \sigma^{(i)}_{tuvw}. \qquad \text{......(17)}$$

Since the Hall conductivity is an axial tensor, the total Hall conductivity tensor is obtained by adding together the contributions from valleys which are related to the ith valley by proper rotations and subtracting the contributions from valleys related to these by inversion or reflection

$$\sigma_{pqr} = \left\{ \pm \sum_j R^{ij}_{pk} R^{ij}_{ql} R^{ij}_{rm} \right\} a^i_{kt} a^i_{lu} a^i_{mv} \sigma^{(i)}_{tuv} \qquad \text{......(18)}$$

(+ sign if $|R^{ij}| = 1$, − sign if $|R^{ij}| = -1$).

Some general results can be derived for the summation of products of elements in the group of symmetry matrices by making use of the theory of groups and their representations. However, for a particular example, the most direct method is to evaluate the sums by inspection. Since many of these sums are zero for any crystal point group, the calculation is greatly simplified by carrying out this summation over valleys first. Further simplification arises from the limited number of non-zero components given by the phenomenological theory, which reduces the number of sums which needs to be evaluated.

When the summations over valleys and over subscripts are completed, the tensor components are expressed in terms of the quantities A, B and C of § 4, the three effective masses m_1, m_2 and m_3 and the direction cosines which define the orientation of the principal axes of the ith valley.

For the crystal class $\bar{3}$m there are several possible valley configurations in **k**-space: (i) one spheroidal valley centred at the origin; (ii) two spheroidal valleys centred on the three-fold axis; (iii) six ellipsoidal valleys centred on the two-fold rotation axes; the axes of the valley centred at $(0, k_y, 0)$ are in the direction of the crystal axes; (iv) six ellipsoidal valleys centred on the reflection planes; the y axis of the valley centred at $(k_x, 0, k_z)$ is in the direction of the y axis of the crystal; (v) twelve ellipsoidal valleys with the ith valley in an arbitrary position and having arbitrary orientation.

The first of these cases is the simple model treated by Bronstein (1932). The second is a trivial extension of this model, which gives identical results. For the twelve-valley model, the components of the conductivity and Hall conductivity tensors are fairly simple but the magnetoconductivity components are so cumbersome that comparison with experiment is very difficult. The two six-valley models are very similar, the first being more restricted in its orientation. In the following section the six-valley model of case (iv) is used and explicit expressions for the independent components of the conductivity tensors are given.

§ 6. THE SIX-VALLEY MODEL

For this model, the symmetry operations relating the valleys can be represented by the following set of matrices:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} C' & -S' & 0 \\ S' & C' & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} C' & S' & 0 \\ -S' & C' & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad \begin{bmatrix} -C' & S' & 0 \\ -S' & -C' & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad \begin{bmatrix} -C' & -S' & 0 \\ S' & -C' & 0 \\ 0 & 0 & -1 \end{bmatrix} \qquad \ldots\ldots(19)$$

in which $C' = \cos\frac{2}{3}\pi$, $S' = \sin\frac{2}{3}\pi$.

The orientation of the ellipsoid with its centre on the (x, z) plane of reflection (the ith ellipsoid) is specified by the direction cosine scheme

$$(a^i_{pq}) = \begin{bmatrix} c & 0 & -s \\ 0 & 1 & 0 \\ s & 0 & c \end{bmatrix} \qquad \ldots\ldots(20)$$

in which $c = \cos\theta$, $s = \sin\theta$, θ being the angle of rotation about the y axis which makes the principal axes of the ith valley coincide with the crystal axes.

The procedure for obtaining the expressions for the overall conductivity tensors will be illustrated by considering the particular component σ_{1122}.

From equation (17), omitting the superscripts ij,

$$\sigma_{1122} = \{\sum R_{1p}R_{1q}R_{2m}R_{2n}\}\, a^i_{pt}a^i_{qu}a^i_{mv}a^i_{nw}\sigma^{(i)}_{tuvw} \qquad \ldots\ldots(21)$$

where the summation in brackets is to be taken over the set of matrices (19). The only non-zero terms in this summation are

$$\sum R_{11}^2R_{22}^2 = \tfrac{9}{4}; \quad \sum R_{11}^2R_{21}^2 = \tfrac{3}{4}; \quad \sum R_{12}^2R_{22}^2 = \tfrac{3}{4}; \quad \sum R_{12}^2R_{21}^2 = \tfrac{9}{4};$$

$$\sum R_{11}R_{12}R_{21}R_{22} = \sum R_{12}R_{11}R_{21}R_{22} = \sum R_{11}R_{12}R_{22}R_{21} = \sum R_{12}R_{11}R_{22}R_{21} = -\tfrac{3}{4}. \qquad \ldots\ldots(22)$$

Each term in (22) is to be taken with the appropriate combination in (21). For example, the particular term $\Sigma R_{11}^{2} R_{22}^{2}$ is associated with

$$a_{1t}^{i} a_{1u}^{i} a_{2v}^{i} a_{2w}^{i} \sigma_{tuvw}^{(i)}. \qquad \ldots\ldots(23)$$

From (20), $a_{2v}=0$ unless $v=2$, $a_{1t}=0$ if $t=2$, and from 8 (*c*) $\sigma_{tuvw}^{(i)}=0$ unless $tuvw$ are equal in pairs. Thus the total contribution from this term is

$$\tfrac{9}{4}(c^2\sigma_{1122}^{(i)} + s^2\sigma_{3322}^{(i)}). \qquad \ldots\ldots(24)$$

Using the explicit expressions for $\sigma_{tuvw}^{(i)}$ given by equation (8 *c*) and summing all contributions, the expression for σ_{1122} given in the list below is obtained.

In this way, it is found that

$$\left.\begin{aligned}
\sigma_{11} &= 3A\left\{\frac{c^2}{m_1} + \frac{1}{m_2} + \frac{s^2}{m_3}\right\} \\
\sigma_{33} &= 6A\left\{\frac{s^2}{m_1} + \frac{c^2}{m_3}\right\} \\
\sigma_{123} &= \frac{-6B}{m_1 m_2 m_3}\left\{s^2 m_1 + c^2 m_3\right\} \\
\sigma_{231} &= \frac{-3B}{m_1 m_2 m_3}\left\{c^2 m_1 + m_2 + s^2 m_3\right\} \\
\sigma_{1111} &= -\frac{3}{2}\frac{C}{m_1 m_2 m_3}\left\{3c^2 s^2\left(\frac{m_3}{m_1} + \frac{m_1}{m_3} - 2\right) + c^2\left(\frac{m_2}{m_1} + \frac{m_1}{m_2} - 2\right)\right. \\
&\qquad\qquad\qquad\qquad\qquad\qquad\left. + s^2\left(\frac{m_2}{m_3} + \frac{m_3}{m_2} - 2\right)\right\} \\
\sigma_{3333} &= \frac{-12C}{m_1 m_2 m_3}\left\{c^2 s^2\left(\frac{m_3}{m_1} + \frac{m_1}{m_3} - 2\right)\right\} \\
\sigma_{1122} &= -\frac{3}{2}\frac{C}{m_1 m_2 m_3}\left\{c^2 s^2\left(\frac{m_3}{m_1} + \frac{m_1}{m_3} - 2\right) + 3c^2\left(\frac{m_2}{m_1} + \frac{m_1}{m_2} - 2\right)\right. \\
&\qquad\qquad\qquad\qquad\qquad\qquad\left. + 3s^2\left(\frac{m_2}{m_3} + \frac{m_3}{m_2} - 2\right) + 8\right\} \\
\sigma_{1133} &= \frac{-6C}{m_1 m_2 m_3}\left\{c^4\frac{m_3}{m_1} + s^4\frac{m_1}{m_3} + c^2\frac{m_3}{m_2} + s^2\frac{m_1}{m_2} + 2c^2 s^2\right\} \\
\sigma_{3311} &= \frac{-6C}{m_1 m_2 m_3}\left\{c^4\frac{m_1}{m_3} + s^4\frac{m_3}{m_1} + c^2\frac{m_2}{m_3} + s^2\frac{m_2}{m_1} + 2c^2 s^2\right\} \\
\sigma_{1123} &= 0 \\
\sigma_{2311} &= 0 \\
\sigma_{2323} &= \frac{6C}{m_1 m_2 m_3}\left\{1 + s^2 c^2\left(\frac{m_3}{m_1} + \frac{m_1}{m_3} - 2\right)\right\}.
\end{aligned}\right\} \ldots(25)$$

It is of interest to consider the particular case $m_1=m_2=m_3$. For this case each valley makes the same contribution to any term and the expressions reduce to those obtained for a single energy minimum and spherical energy surfaces. For example, the Hall coefficient reduces to $-B/6A^2$. From equation (9), with $\lambda=\frac{1}{2}$, this is $-\frac{3}{8}\pi/6n^{(i)}ec$ which is the standard expression for the Hall coefficient of a non-degenerate semiconductor in which the relaxation time is proportional to $E^{-1/2}$, ($6n^{(i)}=N$, the total density of electrons).

Another case of interest is that in which $c=1$, $s=0$. The expressions are then equivalent to those obtained for a model in which there are six valleys with centres on the two-fold axes. For this case, the longitudinal magnetoresistance in the direction of the three-fold axis would vanish.

§ 7. DISCUSSION

The six-valley model of the previous section leads to anisotropy of the resistivity and the Hall coefficient even when only one type of carrier is involved. Also, the three transverse and the two longitudinal magnetoresistance components are generally different from zero. All these predictions are in qualitative agreement with preliminary measurements on both n- and p-type bismuth telluride. The above model is the simplest case which is consistent with these experiments. Further measurements are required to determine whether quantitative agreement can be obtained.

The theory in this paper can be applied with slight modifications to other materials belonging to the crystal class $\bar{3}$m or 3m such as antimony and bismuth. For bismuth, in which a two-band model must be used, a recent analysis of the galvanomagnetic effects has been given by Abeles and Meiboom (1956). They used a three-valley model for the conduction band which is a restricted case of the six-valley model discussed above. Combining this with a single-valley valence band and assuming that the relaxation time of the electrons and of the holes is independent of energy, they obtained satisfactory agreement with their experiments.

REFERENCES

ABELES, B., and MEIBOOM, S., 1954, *Phys. Rev.*, **95,** 31; 1956, *Ibid.*, **101,** 544.
BRONSTEIN, M., 1932, *Phys. Z. Sowjet.*, **2,** 28.
FUMI, F. G., 1952, *Acta Cryst., Camb.*, **5,** 691.
HERRING, C., 1955, *Bell Syst. Tech. J.*, **34** ,237.
JOHNSON, V. A., 1949, *Phys. Rev.*, **75,** 1631.
JONES, H., 1936, *Proc. Roy. Soc.* A, **155,** 653.
JURETSCHKE, H. J., 1955, *Acta Cryst., Camb.*, **8,** 716.
LANGE, P. W., 1939, *Naturwissenschaften*, **27,** 133.

‘Lozenge’ and ‘Tadpole’ Domain Structures on Silicon–Iron Crystals

By L. F. BATES, F.R.S. and P. F. DAVIS

Department of Physics, University of Nottingham

MS. received 4th May 1956 and in final form 13th June 1956

Abstract. New photographs of ‘lozenge’ patterns (Paxton and Nilan) and of ‘tadpole’ patterns (Bates and Hart) on silicon–iron crystals are recorded, and models are given for their interpretation.

---

§ 1. Introduction

In a recent paper Paxton and Nilan (1955) showed photographs of what they termed ‘lozenge’ patterns on a small grain in a thin disc of silicon–iron. They found that these patterns occurred on approximate (110) surfaces which were inclined at an angle of between 4° and 7° to the [001] direction. Their photographs showed domain patterns on the crystal in the demagnetized state and in various stages of magnetization approximately along the [001] direction.

In the demagnetized state, and also in low fields, the lozenges pointing along the [001] direction were arranged in rows parallel to the $[1\bar{1}0]$ direction. When a small field was applied in approximately the [001] direction, alternate rows of lozenge patterns expanded at the expense of neighbouring rows, yet no domain boundaries between the rows of lozenges were visible in the photographs. This may indicate that in low fields there are domains magnetized in [001] and $[00\bar{1}]$ directions separated by walls parallel to the (110) plane, such as are indicated by lines like AA′ in figure 1, which represents the crystal in a demagnetized state. These walls make only a small angle with the surface under investigation, and so a line of contact, such as BB′, of a domain wall with the surface would not be marked by a dense line of colloid, because the field above the region of intersection is very small. The position of a lozenge is indicated in figure 1 by the dotted outline L. At higher field strengths the rows of lozenges ran in approximate $[1\bar{1}1]$ directions, and there appeared groups of what have been described as ‘tadpole’ patterns by Bates and Hart (1953).

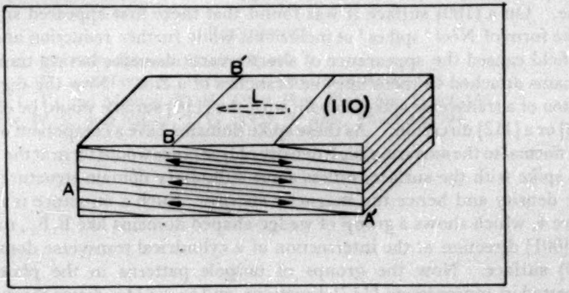


Figure 1.

§ 2. New Experimental Results

The series of photographs, figures 2 and 3 (Plates), which for convenience are turned through 90° with respect to figure 1, show the domain patterns on a similarly orientated grain in a disc of silicon–iron at various stages in a half cycle of a hysteresis loop. Striations in the colloid within the surface of the lozenge domains in these photographs show that the lozenges are magnetized along the easy axis nearest to the surface in [001] or $[00\bar{1}]$ directions.

Figure 2 (*a*) shows part of the surface of the grain when it is saturated by a field parallel to the long edge of the page and no lozenges are visible. In figure 2 (*b*) the field strength has been reduced slightly; vertical lozenge patterns and elongated groups of tadpole structures inclined at 35° to the vertical have appeared. Two of the lozenge patterns are marked by a ring, and a group of tadpoles by a rectangle. Figure 2 (*c*) shows that on further reduction in field the size and number of lozenge patterns have increased, while the number of elongated groups of tadpole patterns has decreased. In figure 2 (*d*) the applied field has been reduced to zero and no tadpole patterns are visible. In figure 2 (*e*) a small reverse field has been applied, causing changes in the sizes of the lozenge patterns. It was not considered worth while to record this process in greater detail.

As the reverse field is increased, figures 2 (*f*), 3 (*a*), (*b*) and (*c*) show that the lozenge patterns become divided into bands, running approximately in $[1\bar{1}1]$ directions; these bands decrease in width as the applied field is increased. The bases of some of the lozenges coincide with dense lines of colloid, which appear as heavy dotted lines, running in approximate $[1\bar{1}1]$ directions. Figures 3 (*a*), (*b*) and (*c*) show also the reappearance and increase in size and number of the elongated groups of tadpole patterns as the field is increased, while figures 3 (*d*), (*e*) and (*f*) show that when the crystal approaches saturation the lozenge patterns are very small, and that the tadpole patterns also shrink until a tsaturation, as in figure 3 (*f*) the material is in a single phase state, the only markings on the surface being the striations which indicate the direction of saturation magnetization in the crystal.

§ 3. Interpretation

A possible interpretation of some of the observed changes in the Bitter patterns with magnetization is as follows. In recent papers (Bates 1955, and Bates and Martin 1956) are described nucleation processes which take place in a crystal magnetized in an easy direction as the applied field is reduced from the saturation value. On a (100) surface it was found that there first appeared small domains in the form of Néel 'spikes' at inclusions, while further reduction and reversal of the field caused the appearance of short reverse domains having transverse spike domains attached to them like the branches of a tree. Now the region of intersection of a transverse spike domain with the (110) surface would be elongated in a $[\bar{1}12]$ or a $[1\bar{1}2]$ direction. As these spike domains have a component of magnetization normal to the surface, a high density of free poles would form at the intersection of a spike with the surface, unless some subsidiary domain structure reduces the pole density and hence the magnetic energy. Such a structure is illustrated in figure 4, which shows a group of wedge-shaped domains like B_1B_1', magnetized in the $[00\bar{1}]$ direction at the intersection of a cylindrical transverse domain with the (110) surface. Now the groups of tadpole patterns in the photographs are elongated in approximate $[1\bar{1}2]$ directions, and so could indicate the intersection of transverse spike domains with the surface. The individual tadpoles in each group are observed to run in $[1\bar{1}1]$ directions, and thus may be narrow bands of free poles between domains such as B_1B_1'. It is therefore thought that the appearance of tadpole patterns as the field is reduced from the saturation value corresponds with the nucleation of Néel spikes at inclusions; and the later appearance of lozenges as the applied field is further reduced corresponds to the formation of the reverse domains studied by Bates and Martin.

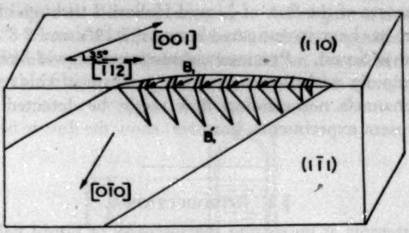


Figure 4.

The rest of the process by which the lozenges become divided into bands running in a $[1\bar{1}1]$ direction is more difficult to interpret, and the process would be greatly influenced by the large demagnetization factor of the crystal in a direction perpendicular to the surface under investigation. Thus, in the demagnetized state the existence of equal volumes of domains magnetized in the $[001]$ and $[00\bar{1}]$ directions is favoured from energy considerations, but in higher fields there comes a stage at which the large demagnetization factor of the thin specimen perpendicular to the observed surface requires that there shall be domains magnetized in $[\bar{1}00]$ and $[0\bar{1}0]$ directions, because the surface is not a true (110) plane. At even greater fields, in the saturated state, the direction of magnetization will be rotated slightly from the $[001]$ direction towards the field which is applied parallel to the surface of the specimen.

ACKNOWLEDGMENTS

Our thanks are due to the Electrical and Allied Industries Research Association for support which made this work possible and for permission to publish it.

REFERENCES

BATES, L. F., 1955, *Research, Lond.*, **8**, 462.
BATES, L. F., and HART, A., 1953, *Proc. Phys. Soc.* A, **66**, 813.
BATES, L. F., and MARTIN, D. H., 1956, *Proc. Phys. Soc.* B, **69**, 145.
PAXTON, W. S., and NILAN, T. G., 1955, *J. Appl. Phys.*, **26**, 994.

The Flow of Liquid Helium I through Narrow Channels

By D. C. CHAMPENEY

Clarendon Laboratory, Oxford

Communicated by K. Mendelssohn; MS. received 15*th May* 1956

Abstract. The nature of the flow of Liquid Helium I through channels of about 1 micron diameter has been investigated between 2·19°K and 2·8°K using pressure heads of up to 3 cm of liquid. Previous measurements in wider channels showed a viscosity falling rapidly with decrease of temperature in this region, suggesting that in narrow channels non-classical flow might be detected even above the λ-point. The present experiments, however, show the flow to be purely classical.

§ 1. Introduction

The early attempts at measuring the viscosity of liquid helium I under its vapour pressure gave conflicting results. Wilhelm, Misener and Clark (1935) used an oscillating cylinder viscometer, Keesom and MacWood (1938) and Keesom and Keesom (1941) measured the damping of an oscillating disc and Johns, Wilhelm and Grayson Smith (1939) used a capillary flow method. However, the more recent measurements by Bowers and Mendelssohn (1950) using capillary flow seem reliable and have been confirmed by de Troyer, van Itterbeek and van den Berg (1951) who employed the oscillating disc method. These determinations show that the viscosity is approximately constant from 4°K down to 3°K, but that it starts to fall rapidly with decrease of temperature below 3°K. This drop, and a similar drop in the entropy just above the λ-point indicate a preparation for the state of He II, and it is of interest to see whether any signs of non-classical flow occur in this temperature region.

Non-classical flow may be expected to show itself in capillary flow by a non-linear dependence of flow rate on pressure head and by a drop in apparent viscosity between 3°K and the λ-point, both effects being more pronounced in narrower channels. Previous work has shown no departure from classical flow, but the narrowest channels from which reliable data could be obtained were of diameter 36μ (Johns, Wilhelm and Grayson Smith) and 40μ (Bowers and Mendelssohn). For a conclusive test, narrower channels of about 1μ diameter are required, and two previous attempts have been made in this laboratory but the results had a large scatter compared with the wide-channel experiments. Bowers and Mendelssohn (1950) used a tube packed with fine wires giving a mean channel diameter of about 1μ, and White (1951) found that the apparatus he used for measuring the flow of He II through etched copper membranes of mean channel diameter about 1μ was unsuited for use above the λ-point giving a large scatter when so used. These observations are shown in figure 3(*a*).

In the present work the flow through a sintered glass filter (such as is used in biological work) has been measured. The mean channel size was stated by the manufacturers to be 1·3μ and the maximum 1·7μ. The scatter in previous work was probably due to partial blockage of the channels by minute bubbles, and precautions have been taken to reduce this effect.

§ 2. METHOD

The filter formed the bottom of a glass reservoir of diameter 2·8 cm which could be raised or lowered in a glass cryostat of diameter 4 cm so as to produce inflow or outflow of liquid from the reservoir. The cryostat formed part of a small Linde helium liquefier, similar to that described by Daunt and Mendelssohn (1948).

The reservoir was surrounded by a copper radiation shield with vertical slits for observing the liquid level in the reservoir. A low power neon lamp provided the illumination. A conical deflector rested on three supports on the bottom of the radiation shield to prevent bubbles formed on the shield from rising and collecting under the filter. The reservoir and shield are shown in figure 1.

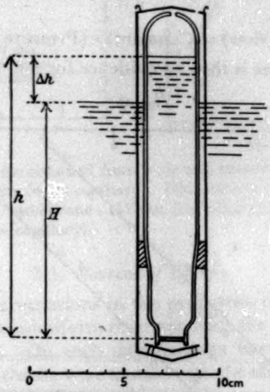


Figure 1. The reservoir and radiation shield.

Measurements were made by either raising or lowering the reservoir, and observing the reservoir and cryostat liquid levels with a cathetometer at half-minute intervals over a period of about 25 minutes. If the rate of change of reservoir level dh/dt was plotted against level difference Δh a straight line was always obtained. However, on some occasions the slope obtained from an inflow differed from that obtained from an outflow, measured immediately afterwards at the same temperature, both giving abnormally slow rates of flow. This was interpreted as being due to the presence of bubbles. Experiments in which inflow and outflow differed by more than 3% have, therefore, been ignored.

It was found that agreement between inflow and outflow could be obtained by keeping the filter at least 13 cm below the mean liquid level, cooling the system below the λ-point and allowing it to warm to the required temperature, stirring the liquid by moving the reservoir up and down and allowing to settle for an hour. Even so, reliable readings were only obtained below 3°K.

The temperature was controlled by a cartesian manostat in the helium pumping line, and the cryostat temperature, as measured on mercury or oil manometers, remained constant to within 2×10^{-3} deg. K at 2·19°K and to within less than

1×10^{-3} deg K at 2·8°K. The absolute value of the temperature was, however, not known to the same accuracy. The usual correction was applied to allow for the hydrostatic pressure of the liquid helium; this amounted to about one hundredth of a degree. However, as there was no visible bubbling the temperature might have been up to a hundredth of degree lower than the values quoted. The 1955 scale was used throughout.

§ 3. RESULTS

3.1. *Pressure Dependence*

Smooth curves were drawn of level height plotted against time and values of dh/dt and Δh obtained. Graphs of $\log_{10} dh/dt$ plotted against $\log_{10} \Delta h$ gave points falling on straight lines of slopes equal to $1 \pm 0{\cdot}05$. Typical curves are shown in figure 2. The full lines are of slope unity. This shows that the flow may be expressed by the relation

$$(\text{Rate of flow}) = (\text{Constant}) \times (\text{Pressure head})^n \qquad \ldots\ldots(1)$$

where $n = 1 \pm 0{\cdot}05$. There is thus no evidence for non-classical flow.

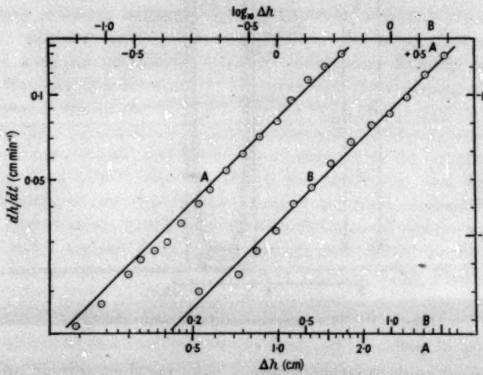


Figure 2. Plot of $\log_{10}(dh/dt)$ against log Δh. A, outflow at 2·20°K; B, inflow at 2·20°K.

Flow below the λ-point was so quick that only three or four readings could be taken before the levels equalized. Reliable values of dh/dt could therefore not be obtained. If a relation of the form of equation (1) is assumed, however, readings taken at 2·16°K and 2·13°K could be approximately represented by a value of n between 0·6 and unity.

3.2. *Temperature Dependence*

The variation of viscosity with temperature was determined from the slopes of graphs of Δh plotted against dh/dt, the viscosity being assumed proportional to $\Delta h/(dh/dt)$. A correction was applied for the variation of density with temperature. For comparison with Bowers and Mendelssohn's curve obtained with a 40μ capillary, the present values have been normalized to 31 μpoise at 2·8°K; the results are shown in figure 3(*b*). The vertical lines connect points obtained from inflow and outflow: neither direction of flow gave results consistently higher than the other. The average deviation of the points from a smooth curve drawn through them is 3%, the smooth curve giving a variation in viscosity between 2·19°K and 2·8°K which is 3% less than that found by Bowers and Mendelssohn. Below the λ-point the value of $\Delta h/(dh/dt)$ at 1 cm level difference was equivalent to viscosities of 1·8μ poise and 1·5 μpoise at 2·16°K and 2·13°K respectively.

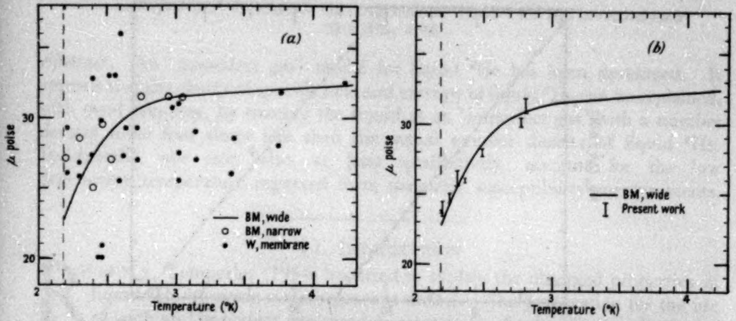


Figure 3. Comparison of results obtained from wide and narrow channels. (*a*) BM, wide : Bowers and Mendelssohn (wide capillary); BM, narrow : Bowers and Mendelssohn (narrow channel); W, membrane : White (membrane). (*b*) BM, wide : Bowers and Mendelssohn (wide capillary).

3.3. *Sources of Error*

Errors could arise from variations in the properties of the filter between one liquefaction and another, non-uniform reservoir bore, the presence of bubbles and the effects of evaporation. On each day readings were taken throughout the temperature range, and all results were treated with the same normalizing factor so that the first above-mentioned effect would give rise to a scatter of results but not to a false temperature dependence. The filter was left pumping in vacuum for a day before each liquefaction. The reservoir bore was uniform to within 1%, such variations occurring over a distance of about 1 cm. This combined with optical flaws in the cryostat and surrounding Dewar flasks could account for the waviness in the curves of log (dh/dt) against log Δh. Again, this will not produce any consistent error. The effects of evaporation from the reservoir may be assumed negligible from the fact that the $[dh/dt, \Delta h]$ graphs give a straight lines through the origin. An alternative method of presenting the results in which the evaporation rate is included explicitly is to fit the values of Δh to the equation

$$\log(\Delta h \pm K_1) = -(1/K_2) \times (\text{time}) \qquad \ldots\ldots(2)$$

This equation may be deduced from the assumption of classical flow (i.e. $n=1$ in equation (1) above) with differing rates of evaporation from reservoir and cryostat. The constant K_1 depends upon the rates of evaporation from the reservoir and from the cryostat, and K_2 is proportional to the viscosity. K_1 had a value between 0·05 cm and 0·06 cm in all cases and was added to Δh for inflow and subtracted from Δh for outflow. This is as expected for a larger rate of evaporation from the cryostat than from the reservoir. The rate of fall of the levels was measured when the reservoir level was K_1 cm above the cryostat level. This rate varied between 0·005 cm min^{-1} and 0·003 cm min^{-1}. From this and the corresponding value of K_1 the reservoir evaporation rate was calculated to be less than 0·001 cm min^{-1}. This correction does not alter the value of *n* and its spread obtained in § 3.1. Typical curves are shown in figure 4. The variation of viscosity with temperature obtained from K_2 agrees with that given in § 3.2.

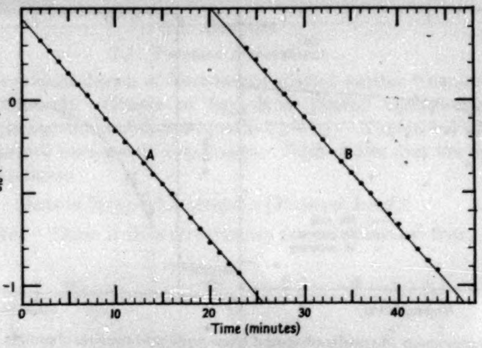


Figure 4. Logarithmic plot of Δh against time.
A, log ($\Delta h+0{\cdot}06$), inflow at 2·20°K; B, log ($\Delta h-0{\cdot}06$), outflow at 2·20°K.

3.4. *Conclusion*

The nature of the flow of He I through narrow channels of about 1 μ diameter is purely classical in the region between 3°K and the λ-point, even though the value of the viscosity starts to fall rapidly with decrease of temperature.

ACKNOWLEDGMENTS

The author wishes to thank Dr. K. Mendelssohn for suggesting this work and giving his advice, and to acknowledge gratefully financial support received from the Department of Scientific and Industrial Research.

REFERENCES

BOWERS, R., and MENDELSSOHN, K., 1950, *Proc. Roy. Soc.* A, **204,** 366.
DAUNT, J. G., and MENDELSSOHN, K., 1948, *J. Sci. Instrum.*, **25,** 318.
DE TROYER, A., VAN ITTERBEEK, A., and VAN DEN BERG, G. J., 1951, *Physica*, **17,** 50.
JOHNS, H. E., WILHELM, J. O., and GRAYSON SMITH, H., 1939, *Canad. J. Res.* A, **17,** 149.
KEESOM, W. H., and KEESOM, P. H., 1941, *Physica*, **8,** 65.
KEESOM, W. H., and MACWOOD, G. E., 1938, *Physica*, **5,** 737.
WHITE, G. K., 1951, *Proc. Phys. Soc.* A, **64,** 554.
WILHELM, J. O., MISENER, A. D., and CLARK, A. R., 1935, *Proc. Roy. Soc.* A, **151,** 342.

The Specific Heat and Entropy of Liquid Helium Three

By S. M. BHAGAT
Department of Physics, University of Delhi

Communicated by F. C. Auluck: MS. received 6*th March* 1956 *and in revised form* 28*th May* 1956

Abstract. An 'imperfect gas' model for liquid ^{3}He has been developed. It appears that the observed specific heat and entropy of liquid ^{3}He can be explained, with good accuracy, by treating the liquid as an 'imperfect gas' with a number density some four times less than the actual number density of liquid ^{3}He. Incidentally, one may also, at least qualitatively, account for the low degeneracy temperature expected from magnetic susceptibility measurements.

§ 1. Introduction

Recently, Temperley (1954) has tried to explain the observed properties of liquid ^{3}He by using a quasi-gaseous model. The justification for the use of such an 'imperfect gas' model seems to be, firstly, the low density of liquid ^{3}He and secondly, the observed viscosity of liquid ^{3}He which is very low. Next, it has been shown (Hammel 1955) that the observed temperature variation of the entropy of liquid ^{3}He can be explained with fairly good accuracy by treating it as an ideal Fermi–Dirac gas with a degeneracy temperature of about 5°K. On the other hand, the recent measurements (Fairbank *et al.* 1954) on the low-temperature variation of the magnetic susceptibility of liquid ^{3}He seem to indicate that the Fermi–Dirac degeneracy temperature used in the entropy calculations is much too high. In fact, it seems to have been established that the observed magnetic susceptibilities can be explained if liquid ^{3}He is treated as a Fermi–Dirac system with a degeneracy temperature an order of magnitude lower than that given by the ideal gas model (i.e. about 0·5°K). It seems, therefore, that there is little justification for relating the ideal Fermi–Dirac gas model to liquid ^{3}He behaviour and a major modification appears to be necessary.

In this connection, we have worked out the specific heat and entropy of liquid ^{3}He in the temperature region 1°K to 2·5°K, treated as an 'imperfect' Boltzmann gas. The present model is essentially similar to that proposed earlier by Temperley (1954). However, keeping in view the susceptibility measurements, we have treated liquid ^{3}He essentially as a classical gas. The imperfections have been introduced by applying corrections of the van der Waals type to the volume and including terms of the type arising from Lennard-Jones–Devonshire '6–12' interactions in the free energy†. It is interesting to note the surprisingly good agreement which one obtains with experiment when the number density in liquid ^{3}He is somewhat modified.

§ 2. The 'Imperfect Gas' Model

As has already been mentioned, the liquid behaviour indicates a much lower degeneracy temperature than the value obtained for an ideal Fermi–Dirac gas with liquid ^{3}He density. Hence it seems reasonable to interpret this lowering as caused by interactions between the ^{3}He atoms. We shall see how, qualitatively, one may expect a lowering of the degeneracy temperature when imperfections are taken into account (increased 'effective' mass and reduced 'effective' number density). In this context it seems worth while to note that the Bose condensation temperature for an ideal Bose–Einstein gas with liquid ^{4}He density is 3·14°K as compared with the observed lambda temperature of liquid ^{4}He, 2·186°K. The low value of the Fermi–Dirac degeneracy temperature obviously permits us to treat liquid ^{3}He, with justification, as a classical Boltzmann system, at least above 1°K.

---

† Here it seems to be of interest to point out that in one of their eariler papers Lennard-Jones and Devonshire (1937) had mentioned that the inclusion of the effect of the finite size of the particles would considerably improve the 'cell' model.

Next, the 'one-particle' (see below) energy – momentum spectrum is taken to be the ideal gas spectrum. This is equivalent to the assumption that the statistical distribution of the energy levels is not affected by the interactions. The justification for this assumption appears to be straightforward. It stems, at least formally, from an argument developed by Feynman (1953). When one considers an assembly of interacting particles the effect of the interactions is to inhibit the motion of each particle by providing a sort of a potential barrier. However, one may treat this inhibition, at least formally, in a somewhat different way as follows: when one atom moves from one position to another (the atoms in the liquid being in a state of kinetic motion) the atoms lying 'along' the line of motion (that is, within the sphere of influence of the moving atom) move out of the way and after the atom has gone through they return to their original positions. Thus one may reduce the final configuration to a form somewhat similar to the original one. Hence, in order to estimate the total energy of any 'particle' in the interacting assembly one may take, instead of the sum of the kinetic and potential energies, the sum of the kinetic energies of the atom itself as well as that of the atoms displaced by it during the course of its motion. In other words, the potential barriers may be considered as replaced by kinetic energy barriers. Next, Feynman (1953) has shown that the displacement of the 'attendant atoms' takes place with roughly the same velocity as that of the atom being studied, so that this extra energy may be taken into account simply by modifying the mass of the 'atoms' of the assembly. He also points out that the situation appears to be somewhat similar to the motion of a solid sphere through a fluid. We have taken the concept of the 'effective mass' from Feynman's theory and have applied it here. However, it is important to note that we have made no attempt to carry out any quantitative investigation into the nature of this concept, which would in fact be very difficult. However, it appears worth while to study the consequences of applying such a conjecture together with some of the more usual methods of taking account of the interactions. With this in mind we write the 'one-particle'† energy spectrum in the form

$$\epsilon(p) = p^2/2m_3^*, \qquad \ldots\ldots(1')$$

where m_3^* is the effective mass of the particle.

To take account of deviations from ideal-gas behaviour one considers the atoms to be of a finite size, by introducing van der Waals corrections in the molar volume. This is equivalent to the assumption that the total volume accessible to an atom is somewhat less than the actual volume of the liquid, and would partly take account of the repulsive part of the interactions between the atoms.

---

† See the discussion concerning the effective number density.

Further, to take account of the potential field between the particles, apart from effects included in the mass modification, we introduce a smoothed-potential term in the free energy. It is usual, when considering liquids (as condensed gases) on the basis of the Lennard-Jones '6–12' potential, to use a 'cell model', in which each atom is treated as moving in a sort of a cell formed by its nearest neighbours (or next nearest neighbours also, as discussed by Hirschfelder *et al.* 1950, 1954). In our model we consider this cell to be formed by all the other atoms in the assembly and the net effect of the other atoms on the atom under consideration is represented by a smoothed-potential function arising from sixth power attractions and twelfth power repulsions. In other words, we add to the free energy of the assembly a term of the form $N_3\chi(N_3V_3^0/N_3)$, with

$$\chi\frac{N_3V_3^0}{N_3} = -\frac{a_3N_3^2}{(N_3V_3^0)^2} + \frac{b_3N_3^4}{(N_3V_3^0)^4}. \qquad \ldots\ldots(1)$$

V_3^0 is the atomic volume and a_3, b_3 are constants to be determined later. It is clear that the form (1) for $\chi(N_3V_3^0/N_3)$ is correct. For instance, Lennard-Jones and Devonshire (1937, 1938), considering the effect of the first shell neighbours, obtain a term of this form in the free energy. Later, Hirschfelder and his collaborators have shown that the inclusion of the effect of the second and third shell does not modify the form of the function but merely alters somewhat the values of the constants. Hence one does not expect the inclusion of a smoothed effect due to all the atoms to change the form of the term. Here it is important to notice one point. It is well known that in liquid ^{3}He the particles have very large zero-point motions. The net effect of these is to give the liquid a volume which is much larger than that expected from the gas kinetic theory. These zero-point motions can therefore be interpreted as very large repulsions (or very small attractions) between the particles in the assembly. In our model no explicit account is taken of the zero-point motions so far but now it seems reasonable to consider $\chi(N_3V_3^0/N_3)$ as the total, zero-point plus potential, energy per atom and this seems to be quite justified since we can still choose our constants a_3 and b_3. It may, of course, be argued that this assumption may not be very correct but then the smoothed-potential model itself is fairly qualitative. Also the introduction of this assumption is not quite so drastic when it is realized that the zero-point motions, after all, are effective merely in increasing the volume of the liquid and giving it its characteristic open structure. They could therefore be accounted for by reducing the value of the constant a_3 in such a way as to give the observed volume.

In short, in the present model liquid ^{3}He is to be treated essentially as a Boltzmann system (for $T > 1°\text{K}$), having a 'one-particle' energy spectrum of the 'ideal-gas' form. Deviations from ideal are accounted for by taking account of the finite size of the atoms and including terms arising from '6–12' interactions in the free energy. Further, both the mass of the particles as well as their number density (see below) are taken to be modified by the presence of interactions. It is not at present quite clear what value should be assigned to the factor by which the number density would be modified and we make no attempt to go into details. However, it is of interest to interpret this factor somewhat as follows.

One should expect that the momentum of particles lying within a certain distance of each other would be rather strongly correlated (cf. Blatt *et al.* 1955). This so-called 'correlation length' would be of the order of the de Broglie wavelength of the particles in the assembly, and in the case of the assembly under consideration (at 1°K) one should expect the latter to be about 10^{-8} cm. Hence, one should expect that the 'correlation volume' would involve only a few particles of the assembly (comparison with experiment below shows that only some four particles are involved). It must be mentioned that the concept of correlated particles must be treated as being characteristically different from that of 'clustered' particles. However, it is clear that here the 'one-particle' energy spectrum refers to all the particles which, lying within the correlation volume, co-operate to give rise to excitations in the assembly, and not to individual ^{3}He atoms.

In this context, it also seems to be of interest to state that in a recent note Mikura (1955) has referred to a suggestion of Morita regarding the treatment of liquid ^{3}He as an ideal Fermi–Dirac gas having a number density of $N_3/4$. However, their considerations do not seem to have led to any further interesting results.

Using the above model we may write the free energy of liquid ^{3}He in the form

$$F_3 = -N_3^*\boldsymbol{k}T + N_3^*\boldsymbol{k}T \ln\left[\frac{N_3^*\boldsymbol{h}^3}{2N_3(V_3^0)_e(2\pi m_3^*\boldsymbol{k}T)^{3/2}}\right] - \left[\frac{a_3N_3^2}{(N_3V_3^0)^2} - \frac{b_3N_3^4}{(N_3V_3^0)^4}\right]N_3^*, \qquad \ldots\ldots(2)$$

with a degeneracy temperature given by

$$\frac{\xi_0}{\boldsymbol{k}} = \frac{\boldsymbol{h}^2}{2m_3^*\boldsymbol{k}}\left(\frac{3N_3^*}{8\pi N_3(V_3^0)_e}\right)^{2/3}, \qquad \ldots\ldots(3)$$

where $N_3(V_3^0)_e = N_3(V_3^0 - v_3^0)$ is the accessible volume, and N_3^* indicates the modified number density. $N_3v_3^0$ here represents the van der Waals correction term.

§ 3. CHOICE OF THE CONSTANTS a_3 AND b_3

We have chosen values for the constants as follows:

$$a_3\boldsymbol{N}^2/\boldsymbol{k} = 6{\cdot}80 \times 10^2, \qquad \ldots\ldots(4)\dagger$$

$$b_3\boldsymbol{N}^4/\boldsymbol{k} = 6{\cdot}29 \times 10^6, \qquad \ldots\ldots(5)\dagger$$

with $\boldsymbol{N} = 6{\cdot}023 \times 10^{23}$ = Avogadro's number, and $\boldsymbol{k} = 1{\cdot}383 \times 10^{-16}$ erg deg^{-1} = Boltzmann's constant.

These values were initially chosen when we had attempted to apply the imperfect-gas model in a somewhat different manner to the study of liquid mixtures of ^{3}He and ^{4}He (Bhagat 1955). In that case it was found that this choice of a_3 and b_3 led to a good qualitative fit with the observed vapour pressure curves. The choice of these values is further justified by the work of Hirschfelder *et al.* (1954), who, in calculating the effect of the second and third shell neighbours in the case of liquid ^{4}He obtain the values (see also Bhagat 1955)

$$a_3\boldsymbol{N}^2/\boldsymbol{k} = 1{\cdot}5 \times 10^4. \qquad \ldots\ldots(4')$$

$$b_3\boldsymbol{N}^4/\boldsymbol{k} = 6{\cdot}3 \times 10^6. \qquad \ldots\ldots(5')$$

We assume here that the values of the constants for ^{3}He and ^{4}He do not differ. Now, if we compare equations (4), (4′) and (5), (5′), respectively, it becomes clear that the choice of b_3 is quite justified. It has been chosen in accord with the more detailed theoretical calculations. The choice of a_3 may at first sight

---

† The choice of these parameters, although quite arbitrary, is fairly critical. Thus slightly different values would completely disturb the agreement with experiment obtained below.

appear to be arbitrary but it becomes quite clear when we recall that we have made the stipulation that a_3 must be chosen in such a way as to include in the $\chi(N_3V_3^0/N_3)$ a contribution from the zero-point motions. In other words, we wish to treat the zero-point motion as an effective decrease in the attractive part of the potential. Having made this stipulation we notice that a_3 and b_3 may be written in terms of the parameters used by Hirschfelder *et al.*, as (see Bhagat 1955) $a_3 = 1{\cdot}2045\Lambda\sigma^6$, $b_3 = 0{\cdot}5055\Lambda\sigma^{12}$.

The parameter σ (having the significance of the low-temperature collision diameter) is obviously very closely related to the specific volume V_3^0. Now, it is required to obtain a specific volume some three times larger than the value given by the classical gas kinetic theory. Hence we must choose a value of σ^3 some three times higher than that taken by Hirschfelder *et al.* Further, since we wish to maintain the same value for b_3 as has been obtained earlier by them, it is natural that we should choose for Λ a value about 81 times lower than that used by them. This would reduce the value of $a_3\boldsymbol{N}^2/\boldsymbol{k}$ used by them by an order of magnitude, that is, we may take the value of $a_3\boldsymbol{N}^2/\boldsymbol{k}$ to be about 10^3 and this does not seem to be very much different from the value of $0{\cdot}68 \times 10^3$ which we have adopted.

§ 4. THE SPECIFIC HEAT

From equation (2) we obtain for the specific heat of liquid ^{3}He the expression

$$C_3 = \frac{3}{2}N_3{}^*\boldsymbol{k} + \frac{N_3{}^*\boldsymbol{k}V_3^0}{V_3^0 - v_3^0}\frac{1}{V_3^0}\frac{\partial V_3^0}{\partial T} + \frac{N_3{}^*\boldsymbol{k}T}{(V_3^0)^2}\left(\frac{\partial V_3^0}{\partial T}\right)^2$$

$$\times\left[-\left(\frac{V_3^0}{V_3^0 - v_3^0}\right)^2 + \frac{6a_3}{\boldsymbol{k}(V_3^0)^4} - \frac{20b_3}{\boldsymbol{k}(V_3^0)^6}\right]$$

$$+\frac{N_3{}^*\boldsymbol{k}T}{V_3^0}\left(\frac{\partial^2 V_3^0}{\partial T^2}\right)\left[\frac{V_3^0}{V_3^0 - v_3^0} - \frac{2a_3}{\boldsymbol{k}(V_3^0)^3} + \frac{4b_3}{\boldsymbol{k}(V_3^0)^5}\right], \qquad \ldots\ldots(6)$$

where

$$V_3^0 = 2(V_3^0)_0[1 + B'T + (1 + C'T^2 + D'T^3)^{1/2}]^{-1} \qquad \ldots\ldots(7)$$

is the empirical expression for the volume (i.e. density) variation recently reported by Kerr (1954). Here $(V_3^0)_0$ is the specific volume at 0°K,

$$B' = 1{\cdot}00559 \times 10^{-3},\ C' = -1{\cdot}95355 \times 10^{-3},\ \text{and}\ D' = -2{\cdot}47787 \times 10^{-2}.$$

Using the values of the constants a_3 and b_3 adopted above and the value of the van der Waals correction term as given by Sommers (1952) we have calculated the value of the molar specific heat of liquid ^{3}He from (6). The values obtained with $N_3{}^* = N_3$ are presented in the first column in table 1. Next, we have normalized these values by making our calculated value for 1°K agree with the value of 1·03 cal deg^{-1} mole^{-1} obtained recently by Roberts and Sydoriak (1954).

Table 1. Molar Specific Heat of Liquid ^{3}He

(1)	(2)	(3)	(4)
1·0	4·30	1·03	1·03
1·2	4·65	1·11	1·15
1·4	5·55	1·32	1·29
1·6	6·25	1·49	1·45
1·8	7·11	1·67	1·64
2·0	8·00	1·91	1·87
2·2	8·95	2·15	2·14
2·4	10·17	2·43	2·44

(1) Temperature (°K); (2) specific heat from eqn (6); (3) normalized value; (4) observed specific heat (Roberts and Sydoriak 1954). (Specific heat in cal deg^{-1}).

The normalization factor turns out to be 4·175. As already stated, this may be interpreted as the factor by which the actual number density in liquid ^{3}He should be reduced. The normalized specific heat values are tabulated in the second column. In the third column we have included the observed values due to Roberts and Sydoriak (1954). It is interesting to mark the rather close agreement between theory and experiment.

§ 5. THE ENTROPY

Expression (2) gives for the entropy of liquid ^{3}He:

$$S=\frac{5}{2}N_3^*\boldsymbol{k}-N_3^*\boldsymbol{k}\ln\frac{N_3^*}{2N_3(V_3^0)_e}\left(\frac{\boldsymbol{h}^2}{2\pi m_3^*\boldsymbol{k}T}\right)^{3/2}+\frac{N_3^*\boldsymbol{k}TV_3^0}{V_3^0-v_3^0}\frac{1}{V_3^0}\frac{\partial V_3^0}{\partial T}$$
$$+\frac{N_3^*\boldsymbol{k}}{V_3^0}\frac{\partial V_3^0}{\partial T}\left[\frac{4b_3}{\boldsymbol{k}(V_3^0)^4}-\frac{2a_3}{\boldsymbol{k}(V_3^0)^2}\right], \qquad \ldots\ldots(8)$$

with $N_3^*=N_3/4{\cdot}175$.

Now, in calculating the entropy it is necessary to fix the value of the degeneracy temperature $\xi_0/\boldsymbol{k}$ given by (3). We have chosen the value 0·45°K because with this value the observed magnetic behaviour of liquid ^{3}He can be explained adequately. The choice of this value implies an effective mass of $6{\cdot}171m_3$. This value seems to be a little high, since Feynman has suggested for liquid ^{4}He an effective mass some three or four times the atomic mass. However, one may expect that due to antiparallel spins the ^{3}He atoms may have somewhat larger attractions and this would lead to a larger number of particles being dragged along. Next, we have calculated the molar entropy and the results are presented in table 2. For comparison we have also included the results obtained from the empirical formula

$$S_1=[3{\cdot}72T-2{\cdot}37T^2+0{\cdot}573T^3+0{\cdot}141T^4-0{\cdot}053T^5]\ \text{cal deg}^{-1}\,\text{mole}^{-1}$$

for $0{\cdot}5\leqslant T\leqslant 2{\cdot}25$°K, given recently by Abraham *et al.* (1954). The agreement between theory and experiment is again surprisingly good.

Table 2. Molar Entropy of Liquid ^{3}He

(1)	(2)	(3)	(1)	(2)	(3)
1·0	2·05	2·01	2·0	3·08	3·10
1·2	2·24	2·21	2·2	3·32	3·38
1·4	2·44	2·39	2·4	3·64	3·66
1·6	2·64	2·60	2·6	3·88	3·87
1·8	2·86	2·84			

(1) Temperature (°K); (2) entropy, calculated; (3) entropy, observed (Abraham *et al.* 1954). (Entropy in cal deg^{-1}.)

ACKNOWLEDGMENTS

In conclusion, it is a pleasure to express my deep sense of gratitude to Professors D. S. Kothari and F. C. Auluck for their kind interest and constant encouragement throughout the course of this work. I am also thankful to Dr. V. S. Nanda and Mr. R. K. Pathria for several helpful discussions.

REFERENCES

ABRAHAM, B. M., OSBORNE, D. W., and WEINSTOCK, B., 1954, *Phys. Rev.*, **94**, 202.
BHAGAT, S. M., 1955, *Ph.D. Dissertation*, pp. 163–165, University of Delhi.

Blatt, J. M., Butler, S. T., and Schafroth, M. R., 1955, *Phys. Rev.*, **100**, 481.

Fairbank, W. M., Ard, W. B., and Walters, G. K., 1954, *Phys. Rev.*, **95**, 566.

Feynman, R. P., 1953, *Phys. Rev.*, **91**, 1291, 1301.

Hammel, E. F., 1955, Paper on ' Properties of ^{3}He at Low Temperatures ' in *Progress in Low Temperature Physics*, Vol. 1 (Amsterdam : North-Holland Publishing Co.). Ed.: C. J. Gorter.

Hirschfelder, J. O., Wentorf, R. H., and Buehler, R. J., 1950, *J. Chem. Phys.*, **18**, 1484.

Hirschfelder, J. O., Wentorf, R. H., and Curtiss, C. F., 1954, *Molecular Theory of Liquids and Gases* (London and New York : Wiley) p. 293.

Kerr, E. C., 1954, *Phys. Rev.*, **96**, 552.

Lennard-Jones, J. E., and Devonshire, A. F., 1937, *Proc. Roy. Soc.* A, **163**, 53; 1938, *Ibid*, **165**, 1.

Mikura, Z., 1955, *Progr. Theor. Phys.*, *Japan*, **13**, 120.

Roberts, T. R., and Sydoriak, S. G., 1954, *Phys. Rev.*, **93**, 1418.

Sommers, H. S., 1952, *Phys. Rev.*, **88**, 113.

Temperley, H. N. V., 1954, *Proc. Phys. Soc.* A, **67**, 495.

The Form of the Effective Electronic Potential in a Crystal

By G. G. HALL†

Department of Theoretical Chemistry, University of Cambridge

MS. received 4th July 1955 and in revised form 4th July 1956

Abstract. The equations for the best single-electron orbitals in a crystal are interpreted in terms of an effective potential and the forms of the various contributions to this potential are considered. The exchange terms are important and act like an attractive potential which moves through the crystal. The form of this attraction changes abruptly and its magnitude decreases greatly on passing from orbitals arising from the valence shell to higher orbitals. An approximate form for the exchange potential, suggested recently by Slater, is criticized in the light of these results.

§ 1. INTRODUCTION

MOST methods of calculating single-electron wave functions for solids are based on the solution of an equation of the Schrödinger type

$$(-\tfrac{1}{2}\nabla^2 + V')\psi = E\psi.$$

(Throughout this paper Hartree atomic units will be used and the orbitals will be functions only of position.) The potential V' used in this equation includes the attractive potential of the nuclei and the repulsive potential of the other electrons and is usually tabulated numerically. This repulsive potential is defined by the Hartree–Fock equations and involves all the other occupied orbitals. It ought to be determined by an iterative procedure using all the occupied orbitals but this has not yet proved practicable for solids.

Related to this problem of determining the potential is the problem of determining the general form of the orbitals themselves, particularly the unoccupied orbitals. This includes the problem of the significance of these orbitals and of their relation to the extra orbitals occupied in ions and excited states.

In this paper the form of the potential in a simple example is first considered so that the main features can be displayed. The form is then discussed in general using approximate orbitals.

§ 2. THE HYDROGEN ATOM

The simplest example of a potential for an orbital is in the hydrogen atom. The Hartree–Fock equations are first derived for an atom having N singly occupied orbitals with parallel spins and then specialized for hydrogen by putting N equal to 1. This is rather an artificial procedure and may tell us nothing new about hydrogen but, since it preserves the form of the orbital equations, it may be used to study these equations. The resulting orbital equation is

$$\left\{-\tfrac{1}{2}\nabla_1^2 - \frac{1}{r_1} + \int \frac{\psi_0^2(2)}{r_{12}}\, dV_2\right\}\psi(1) - \psi_0(1)\int \frac{\psi_0(2)\,\psi(2)}{r_{12}}\, dV_2 = E\psi(1),$$

where ψ_0 is the occupied orbital, the subscript and the argument 1 denote the spatial coordinates of the electron and the integrations are over the space of the vector 2 only. When $\psi=\psi_0$ this equation reduces to the familiar Schrödinger equation, as is obviously necessary, so that

---

† Now at Department of Mathematics, Imperial College, London S.W.7.

$$V'=-1/r_1, \qquad \psi_0(1)=(1/\sqrt{\pi})\exp(-r_1).$$

The orbital equation can now be written explicitly as

$$\left\{-\tfrac{1}{2}\nabla_1^2-\left(1+\frac{1}{r_1}\right)\exp(-2r_1)\right\}\psi(1)-\frac{1}{\pi}\exp(-r_1)\int\frac{\exp(-r_2)\psi(2)}{r_{12}}\,dV_2=E\psi(1).$$

In this form the electron–nuclear attraction is cancelled by part of the first integral. The solutions of this equation other than ψ_0 are clearly very different from excited hydrogen wave functions for the potential is quite different from the simple Coulomb one and consists of a short-range attraction and an exchange interaction between the electron in the occupied orbital and what may be called a virtual electron in the orbital being considered. Such orbitals are known as virtual orbitals. They are the same as the orbitals of a test charge introduced into the system since the interactions appear as for an additional electron but the original charge distribution is not disturbed. This can be verified by observing that the equation becomes identical with that for the outer orbitals of the hydrogen negative ion in triplet excited states with electron spins parallel provided that the inner orbital of the ion is assumed to remain ψ_0. For the lower triplet states of this ion this inner orbital is unlikely to be close to ψ_0.

Thus the potential given by the Hartree–Fock equation is very different in form when occupied and virtual orbitals are considered. For the occupied orbital there is a strong Coulomb attractive potential which will concentrate the orbital around the nucleus. For the virtual orbital there is only a weak short-range penetration potential and an exchange interaction which is also weak because it depends on an integral over orthogonal orbitals. The virtual orbitals, therefore, are probably diffuse unbound states. Although they are solutions of the same eigenvalue equation, the physical significance of the virtual orbitals is quite different from that of the occupied orbital and can be clarified only by introducing the idea of a test charge.

§ 3. CONTRIBUTIONS TO THE POTENTIAL

The potential energy in an orbital equation for a molecule or a solid is due, according to the Hartree–Fock equations, to the nuclear attractions and to the averaged electronic repulsions. The nuclear attractions contribute a term

$$-\sum_{\alpha}(Z_\alpha/r_{1\alpha}),$$

where α refers to a nucleus. The electronic repulsions contribute two terms one of which, if all the occupied orbitals are doubly occupied, is

$$V(1)=\int\frac{\rho(2,2)}{r_{12}}\,dV_2$$

with

$$\rho(2,1)=2\sum_n\bar{\psi}_n(2)\psi_n(1).$$

This summation is over the space orbitals ψ_n and the factor 2 allows for their double occupation. The second term is the exchange potential A defined by

$$A(1)\psi(1)=-\tfrac{1}{2}\int\frac{\rho(2,1)\psi(2)}{r_{12}}\,dV_2.$$

If some of the orbitals are singly occupied these definitions will require modification. The nuclear attractions and V each diverge at all points as the number of particles in the system increases but their sum converges.

The density matrix $\rho(2, 1)$ is invariant to unitary transformations of the orbitals (Hall and Lennard-Jones 1950). It is convenient, therefore, to consider it as

$$\rho(2, 1) = 2 \sum_n \bar{\chi}_n(2)\, \chi_n(1),$$

where the orbitals χ_n are equivalent orbitals. These orbitals are solutions of the self-consistent equations and may be defined by the condition that their self-energy

$$\sum_n \int \frac{\bar{\chi}_n(1)\, \chi_n(1)\, \bar{\chi}_n(2)\, \chi_n(2)}{r_{12}}\, dV_1\, dV_2$$

is a maximum with respect to unitary transformations of occupied orbitals. Since the self-energy is a measure of localization these equivalent orbitals will be the most localized orbitals. It has been shown (Bhattacharya 1953) that, under normal circumstances, orbitals which permute into one another under the operations of a symmetry group will be equivalent orbitals whereas the extended crystal orbitals are the least localized orbitals. By transforming to localized orbitals the difficulty of normalizing the delocalized crystal orbitals is avoided. This definition also enables us to relate these equations to the original Hartree equations which have no exchange terms. The Coulombic and exchange parts of the total energy are each invariant to transformations of orbitals but their common part is the self-energy and this is not invariant. The remaining part of the exchange energy will be at a minimum for equivalent orbitals and so for these only will a treatment without exchange be possible. The Hartree equations will have the same form as those discussed above except that in place of the exchange operator A there will be an operator B_n arising from the cancelling self-energy where

$$B_n(1)\, \chi_n(1) = -\tfrac{1}{2} \int \frac{\bar{\chi}_n(2)\, \chi_n(2)}{r_{12}}\, dV_2\, \chi_n(1).$$

§ 4. POTENTIALS FOR IONIC CRYSTALS

For an ionic solid all the χ_n are localized around single nuclei. The electron density close to any one nucleus is almost entirely due to the electrons in orbitals localized around that nucleus. The V term in the potential is therefore the sum of the corresponding terms in the potentials for the isolated ions, provided that the χ_n are little changed. Since these χ_n are orthonormal they are not identical with those of the free ions, but the difference should not be significant at ordinary internuclear distances.

In order to examine in more detail the form of the potential it is convenient to consider the integral

$$I_{nt}(1) = \int \frac{\chi_n(2)\, \chi_t(2)}{r_{12}}\, dV_2.$$

If the nucleus associated with χ_n is taken as origin, $1/r_{12}$ can be expanded in spherical harmonics around it giving (cf. Eyring, Walter and Kimball 1944, p. 371)

$$\frac{1}{r_{12}} = \sum_l \left(\frac{r_<{}^l}{r_>{}^{l+1}}\right) \sum_m \left[\frac{(l-|m|)\,!}{(l+|m|)\,!}\right] \mathrm{P}_l{}^m(\cos\theta_1)\, \mathrm{P}_l{}^m(\cos\theta_2) \exp[im(\omega_1-\omega_2)].$$

Thus I_{nt} becomes

$$\begin{aligned} I_{nt} &= \sum_{lm} \left[\frac{(l-|m|)\,!}{(l+|m|)\,!}\right] \int \chi_n \chi_t \left(\frac{r_<{}^l}{r_>{}^{l+1}}\right) \mathrm{P}_l{}^m(1)\, \mathrm{P}_l{}^m(2) \exp[im(\omega_1-\omega_2)]\, dV_2 \\ &= \int \left(\frac{\chi_n \chi_t}{r_>}\right) dV_2 + \tfrac{1}{2}\mathrm{P}_1{}^1(1) \exp(i\omega_1) \int \chi_n \chi_t \frac{r_<}{r_>{}^2}\, \mathrm{P}_1{}^1(2) \exp(-i\omega_2)\, dV_2 + \dots . \end{aligned}$$

Each of these integrals can be split into two parts depending on which of r_1 or r_2 is the greater. It is convenient instead to divide them so that the one part has $r_> = r_1$, $r_< = r_2$ and is integrated over all r_2 while the other part is integrated from 0 to r_1 and includes terms which cancel portions of the first part. The first part then gives the complete asymptotic expansion of the potential while the second part gives the penetration potential which modifies the asymptotic form at small distances. The asymptotic form of the integral is then

$$I_{nt} \sim \int \frac{\chi_n \chi_t}{r_1} dV_2 + \mathbf{r}_1 \cdot \int \frac{\chi_n \chi_t \mathbf{r}_2}{r_1^{\,3}} dV_2 + \dots$$

$$\sim \frac{\delta_{nt}}{r_1} + \mathbf{r}_1 \cdot \frac{\mathbf{M}_{nt}}{r_1^{\,3}} + \dots ,$$

where $\mathbf{M}_{nt}$ is the transition moment between the orbitals. The penetration terms also have this $1/r$ form for small r and so counteract the asymptotic terms there but they vanish for large r. Their form, for Slater atomic orbitals, is given by Roothaan (1951).

The potential V at any point is given by the sum of the potentials due to orbitals around the nearest nucleus and to other orbitals. If the orbitals are assumed to be unaltered, the first of these potentials is the same as in an isolated ion. The penetration potentials around other nuclei, because of their rapid decrease with distance, can be ignored except possibly those of the nearest negative ions. The asymptotic terms of these orbitals are still important. The first part of the asymptotic expansion combines with the electron–nuclear attractions to give a Madelung potential and so eliminates the two divergent terms in the potential. The second part of the expansion depends on the dipole moment of the orbitals. The symmetry of the orbitals will usually force these terms either to vanish or cancel. The quadrupole terms do not vanish but they decrease fairly quickly with distance so that only the nearest need be considered.

The form of the potential A is less obvious. It is only defined in relation to the orbital on which it acts. If this orbital is occupied then in general it can be written as

$$\psi = \sum_s T_s \chi_s,$$

where χ_s are the occupied equivalent orbitals. This leads to

$$A\psi = -\sum_{ns} \chi_n T_s I_{ns}$$

the asymptotic form of which is

$$A\psi(1) \sim -\sum_s (T_s \chi_s / r_s) - \sum_{ns} (T_s \chi_n \mathbf{r}_s . \mathbf{M}_{ns} / r_s^{\,3}) - \cdots,$$

where r_s is the distance from the nucleus s to the field point 1. The first term here predominates and the second term frequently will vanish. When 1 is closest to a nucleus t the major term behaves approximately as the potential of a positive charge on t, viz.

$$A\psi \simeq -(\sum_s T_s \chi_s)/r_t = -(1/r_t)\psi,$$

but, if 1 moves closer to another nucleus, A turns into an attraction towards that nucleus. Thus the potential A can be interpreted as the potential of a positive hole whose location varies with the point being considered and whose effect is that of a unit positive charge on the nearest nucleus. This gives a more precise picture of the moving Fermi hole discussed by Wigner and Seitz (1934). This approximation to the potential becomes even more accurate in the space between two nuclei though there is a smooth transition from one centre of attraction to the next. The penetration part of the potential A behaves in the same way as the asymptotic part but its form depends on the orbitals. Its effect will be to cancel the $1/r$ attraction at very small distances from the nucleus and to vanish rapidly for larger r. Thus the exchange potential in the neighbourhood of an ion behaves very much like the exchange potential of the free ion. The crystal potential in an ionic crystal is therefore very similar to that acting on the corresponding bound electron in the nearest free ion but with the addition of the Madelung potential and the penetration potentials of neighbouring ions.

The effect of the exchange potential A on an unoccupied orbital is quite different from its effect on an occupied orbital. Since the unoccupied orbitals are orthogonal to all the occupied ones, the first part in the asymptotic expansion always vanishes. There is therefore no moving potential hole. If strong dipole transitions are possible between occupied and unoccupied orbitals, there will be a dipole potential. Other contributions to the potential will make it the same as for the virtual orbitals of the free ion and not their excited orbitals, except that there will be the Madelung field. This potential is much more nearly uniform than for the occupied orbitals, and the unoccupied orbitals will be close to plane waves. Thus the discontinuity in the effect of the exchange potential occurs both in the free ions and in the crystal.

The equivalent orbitals χ_s are determined by equations containing the same potential energy terms as those above. The exchange term A acts as an attractive potential centred on the fixed nucleus s and with the asymptotic form $A\chi_s \sim -(1/r_s)\chi_s$. The potentials are therefore very close to those determining the orbitals of the free ions except that the Madelung field will decrease the potential around a positive ion and increase it around a negative ion. A fixed and localized potential like this will lead to localized orbitals so that it is consistent to take these orbitals to be localized. The Hartree equations with the potential B_s will give almost the same result since B_s will be the major term in the potential A.

§ 5. POTENTIALS FOR VALENCE CRYSTALS

In a valence crystal some of the equivalent orbitals are localized around two centres and this makes it more difficult to discuss the potentials accurately. Most of the results found above still hold but in a generalized form.

From the previous discussion it is clear that the form of the potential will depend primarily on the integral I_{nn} for a two-centre orbital so this must first be found. The two-centre orbital χ_n may be expressed very approximately in terms of the two atomic orbitals as

$$\chi_n(2) = [\lambda\omega_a(2) + \mu\omega_b(2)]/\sqrt{(\lambda^2 + \mu^2 + 2\lambda\mu S)},$$

where ω_a and ω_b are normalized atomic orbitals around the nuclei a and b, and

$$S = \int \omega_a \omega_b \, dV.$$

The integral I_{nn} then becomes

$$\int \frac{\chi_n^2(2)}{r_{12}} dV_2 = \left[\lambda^2 \int \frac{\omega_a^2(2)}{r_{12}} dV_2 + \mu^2 \int \frac{\omega_b^2(2)}{r_{12}} dV_2 + 2\lambda\mu \int \frac{\omega_a(2)\,\omega_b(2)}{r_{12}} dV_2\right] \Big/ (\lambda^2 + \mu^2 + 2\lambda\mu S).$$

The last integral here is a two-centre one but it may be estimated in terms of one-centre integrals by using the Mulliken approximation viz:

$$\int \frac{\omega_a \omega_b}{r_{12}}\, dV_2 \simeq \tfrac{1}{2} S \left[\int \frac{\omega_a^2(2)}{r_{12}}\, dV_2 + \int \frac{\omega_b^2(2)}{r_{12}}\, dV_2 \right].$$

The accuracy of this approximation has recently been discussed by Ellison (1955). This gives

$$I_{nn} \simeq [(\lambda^2 + \lambda\mu S)/(\lambda^2 + \mu^2 + 2\lambda\mu S)] \int \frac{\omega_a^2(2)}{r_{12}}\, dV_2 + [(\mu^2 + \lambda\mu S)/(\lambda^2 + \mu^2 + 2\lambda\mu S)] \int \frac{\omega_b^2(2)}{r_{12}}\, dV_2.$$

The separation of these single-centre integrals into asymptotic and penetration parts can then be carried out as before.

In the orbital equation the long range nuclear attractions combine with the asymptotic part of the V potential to give a Madelung field whose charges depend on the parameters λ and μ. For a homonuclear crystal $\lambda = \mu = 1$ so that

$$I_{nn} \simeq \tfrac{1}{2} \int \frac{\omega_a^2}{r_{12}}\, dV_2 + \tfrac{1}{2} \int \frac{\omega_b^2}{r_{12}}\, dV_2$$

and this Madelung field vanishes. The Coulomb part of the potential is then just the sum of the Coulomb potentials of the free atoms and, to give the same terms as in I_{nn}, the valence electrons have to be taken in singly occupied orbitals directed along the bonds. The allocation of spins to these valence electrons in the atom makes little difference to the potential but it is convenient to take them all parallel so that the orbitals can be transformed freely. The exchange potential in the crystal has the approximate asymptotic form, in the neighbourhood of the bond ab,

$$A\psi \sim -\tfrac{1}{2} \sum_s \left(\frac{1}{r_{as}} + \frac{1}{r_{bs}} \right) T_s \chi_s \simeq -\tfrac{1}{2} \left(\frac{1}{r_a} + \frac{1}{r_b} \right) \psi.$$

To give the exchange hole of the atom the correct asymptotic charge and shape for the solid, additional potentials of $1/2r$ on the central atom and $-1/2r$ on each nearest neighbour are required. The penetration part of the exchange potential will modify these potentials close to the nuclei so that around the central atom it will probably be sufficient to ignore the additional potential on that atom and include only the $-1/2r$ on its neighbours.

The integral I_{nt} for orbitals χ_n and χ_t in a valence crystal is generally very small because of the localization of the orbitals χ_n and χ_t in different parts of space. This holds even if the orbitals extend over a common atom provided that the atomic orbitals involved are localized in different directions or have different total quantum numbers. The one exception will occur when χ_t involves the same directed atomic orbitals, and this it does when χ_t is the anti-bonding orbital

$$\chi_t = [\mu\omega_a - \lambda\omega_b]/\sqrt{(\lambda^2 + \mu^2 - 2\lambda\mu S)}.$$

For this, in the same way as before,

$$I_{nt} \simeq [(\lambda\mu + \tfrac{1}{2}S\mu^2 - \tfrac{1}{2}S\lambda^2)/\sqrt{(\lambda^4 + 2\lambda^2\mu^2 + \mu^4 - 4\lambda^2\mu^2 S^2)}] \int \frac{\omega_a^2}{r_{12}}\, dV_2 + [(-\lambda\mu + \tfrac{1}{2}S\mu^2 - \tfrac{1}{2}S\lambda^2)/\sqrt{(\lambda^4 + 2\lambda^2\mu^2 + \mu^4 - 4\lambda^2\mu^2 S^2)}] \int \frac{\omega_b^2}{r_{12}}\, dV_2,$$

and if $\lambda = \mu = 1$

$$I_{nt} \simeq \tfrac{1}{2}(1 - S^2)^{-1/2} \int \frac{\omega_a^2}{r_{12}}\, dV_2 - \tfrac{1}{2}(1 - S^2)^{-1/2} \int \frac{\omega_b^2}{r_{12}}\, dV_2.$$

This is almost the same as I_{nn} so that the exchange operator, when acting on these unoccupied anti-bonding orbitals, will be equivalent to the same potential as for the occupied bonding orbitals. The change in the character of the potential will occur not between occupied and unoccupied orbitals but between orbitals arising from the valence shell and those in higher shells.

If the two-centre orbitals are not as localized as the approximation above implies the main conclusions are not changed. The range of the penetration potential would have to be increased and the charges in the asymptotic potential decreased as the localization of the orbital decreased. Nevertheless it is consistent to assume that χ_n is localized to two atoms since this leads to a potential, in the equation determining χ_n, which is also concentrated around two nuclei. The localization of an anti-bonding orbital χ_t is clearly less than that of χ_n since, although the potential is very similar to that for χ_n, the orbital is less stable.

§ 6. POTENTIALS FOR A METAL

The discussion of the form of the potential for a metal is made more complicated by the fact that no transformation of the occupied orbitals to orbitals localized around one or two nuclei is possible. Some features of the potential may, however, be investigated by considering solids, such as the alkali metals, for which there are localized orbitals in the standard excited state (Hall 1953). The crystal orbitals may still be expressed as

$$\psi_n = \sum_s T_{ns}\,\chi_s,$$

where χ_s are localized around single atoms, but there will usually be many more χ_s than occupied crystal orbitals, so that unitary transformations are no longer possible. These orbitals χ_s cannot be defined in terms of the ground state of the crystal alone. They can be defined as the localized equivalent orbitals of the standard excited state or, more approximately, as atomic orbitals orthogonalized by Löwdin's procedure (1950).

The electron density matrix is now

$$\rho(2,\,1) = 2\sum_{\text{occ}} \bar{\psi}_n(2)\,\psi_n(1) = \sum_{uv} p_{uv}\,\bar{\chi}_u(2)\,\chi_v(1)\,; \qquad p_{uv} = 2\sum_{\text{occ}} \bar{T}_{nu}\,T_{nv}.$$

On physical grounds it is reasonable to expect the charge to be uniformly spread in a monatomic crystal so that, if there are as many electrons as equivalent orbitals per atom, $p_{uu}=1$. The potential $V(1)$ now has two parts

$$V(1) = \sum_u \int \frac{\bar{\chi}_u(2)\,\chi_u(2)}{r_{12}}\,dV_2 + \sum_{u\neq v} p_{uv} \int \frac{\bar{\chi}_u(2)\,\chi_v(2)}{r_{12}}\,dV_2,$$

and the asymptotic expansion of the first part yields the $1/r_\alpha$ terms which cancel the electron–nuclear attractions. The second part is very small because of the localization of χ_u and χ_v in different parts of space. This means that V is just the sum of the corresponding parts of the potentials of the free atoms.

The exchange potential is written

$$A\psi_n(1) = -\tfrac{1}{2}\sum_s T_{ns} \sum_{uv} p_{uv}\,\chi_v(1) \int \frac{\chi_u(2)\,\chi_s(2)}{r_{12}}\,dV_2$$

and, if χ_s is localized around the nucleus s, the asymptotic form of this is

$$\begin{aligned} A\psi_n &\sim -\tfrac{1}{2}\sum_s T_{ns} \sum_v \frac{p_{sv}\,\chi_v(1)}{r_{s1}} - \cdots \\ &\sim -\tfrac{1}{2}\sum_s \frac{T_{ns}\,\chi_s(1)}{r_{s1}} - \tfrac{1}{2}\sum_s T_{ns} {\sum_v}' \frac{p_{sv}\,\chi_v(1)}{r_{s1}} - \cdots . \end{aligned}$$

The first of these terms has the form of the moving hole which has been examined above except that the charge is now only $\frac{1}{2}$. The second term is not negligible, however, but corresponds to the charge

$$\tfrac{1}{2}\sum_{v}{}' p_{sv}\chi_v(1)/\chi_s(1),$$

the summation being over all values of v not equal to s, so that, at the mid-point between nuclei, the total charge is $\frac{1}{2}(1+p_{sv})$ and this will probably be close to unity. Thus the hole has the potential of a charge of $\frac{1}{2}$ close to the nucleus but of almost 1 between neighbouring nuclei. This reduction of the charge will be counteracted to a large extent by the penetration part of the exchange potential which is multiplied by the same factor and which cancels the asymptotic part close to the nucleus. The final result is that the potential around an atom for a crystal orbital is very close to that for the corresponding atomic orbital.

This result for the potential does not depend on whether ψ_n is occupied or unoccupied, so long as the same set of orbitals χ_s is used to describe it. When ψ_n is more highly excited than can be described in this way, the exchange potential will become very much smaller so that there will be a break in the form of the potential. In general then the break in potential seems to occur between orbitals continuous with valence shell atomic orbitals, and orbitals continuous with virtual atomic orbitals, rather than between occupied and unoccupied orbitals as for ionic crystals. It is interesting to note that the free-electron treatment of metals does not give this result. In it the exchange potential (cf. Seitz 1940, p. 340) falls smoothly but rapidly in value on passing from occupied to unoccupied orbitals. This is probably an illustration of the limitations of the free-electron method. Nevertheless it seems reasonable that there should be some reduction of the potential on passing from occupied to unoccupied orbitals in metals as in valence crystals. This reduction is found in molecules and makes the true virtual orbitals more diffuse than the approximate virtual orbitals calculated using localized orbitals. For this very reason the approximate orbitals are of more value in describing excited states of molecules.

§ 7. PHYSICAL SIGNIFICANCE OF VIRTUAL ORBITALS

Virtual orbitals were defined as solutions of the self-consistent field equations which are unoccupied. Their physical significance is seen only when these equations are related to those for the solid when charged or excited. In the negative ion of a crystal, for example, the extra electron is distributed over the whole crystal so that the disturbance of the potential field at any point, and hence of the occupied orbitals, will be negligible. (This does not follow for a smaller system such as an atom or a molecule.) The equation for the orbitals is

$$(H+V+A+V_x+A_x)\psi=E\psi,$$

where V and A have the same form as for the neutral crystal and

$$V_x(1)=\int\frac{\psi_x^2(2)}{r_{12}}\,dV_2:\;\; A_x\psi=-\psi_x(1)\int\frac{\bar{\psi}_x(2)\psi(2)}{r_{12}}\,dV_2$$

The extra orbital ψ_x is a crystal orbital spread over the whole crystal and normalized to unity so that V_x will be very small and A_x at least as small. For $\psi=\psi_x$ these two terms exactly cancel so that ψ_x is just a virtual orbital of the neutral crystal. The various virtual orbitals are therefore the various orbitals into which the extra electron in the negative ion can be put.

When an electron is excited from one orbital to another there are several terms of form similar to V_x and A_x. If the system is large enough the effect of these extra terms is negligible and the orbitals are the same as those of the crystal in its ground state. In particular the excited orbital is just a virtual orbital. If the system is an atom or a molecule, on the other hand, the extra terms are significant. The excited orbital is most affected and becomes more firmly bound than a virtual orbital.

§ 8. DISCUSSION

The general result of these investigations is that the crystal potential in most crystals should be quite close to the superposition of the fields of the free atoms or ions. The valence electrons of diamond, for example, have a potential equal to that for an electron in an outer orbital of a carbon atom in its ^5S state with extra potentials of $1/2r$ and a penetration term which counteracts it for small r and also $-1/2r$ attractions on its four nearest neighbours. This potential is independent of orbital for the bands which arise from the valence shell of the atom but for higher unoccupied bands the potential is just the superposition of the potentials for the virtual orbitals of the carbon atom.

Slater (1951) has proposed that the exchange potential should be represented approximately by a potential proportional to the cube root of the electron density. This particular form was suggested by a study of the free-electron model. It does not agree very well with the results above for it suggests that the potential is greatest at the nuclei and falls away approximately exponentially whereas, because of its penetration part, the potential is very small at the nuclei and falls away as $1/r$. Furthermore it is incorrect, according to both the analysis above and the free-electron results, to use this approximate form for all the unoccupied orbitals.

ACKNOWLEDGMENT

Most of this work was done while the author was a Vacation Consultant at the Radar Research Establishment, Malvern, and his thanks are due to Dr. L. Pincherle and his colleagues for their hospitality and many useful comments.

REFERENCES

BHATTACHARYA, B., 1953, *Ph. D. Dissertation*, University of Cambridge.
ELLISON, F. O., 1955, *J. Chem. Phys.*, **23**, 2358.
EYRING, H., WALTER, J., and KIMBALL, G. E., 1944, *Quantum Chemistry* (New York: John Wiley).
HALL, G. G., 1953, *Proc. Phys. Soc.* A, **66**, 1162.
HALL, G. G., and LENNARD-JONES, J. E., 1950, *Proc. Roy. Soc.* A, **202**, 155.
LÖWDIN, P. O., 1950, *J. Chem. Phys.*, **18**, 365.
ROOTHAAN, C. C. J., 1951, *J. Chem. Phys.*, **19**, 1445.
SEITZ, F., 1940, *Modern Theory of Solids* (New York: McGraw-Hill).
SLATER, J. C., 1951, *Phys. Rev.*, **81**, 385.
WIGNER, E., and SEITZ, F., 1934, *Phys. Rev.*, **46**, 509.

A Theory of Magnetic Double Refraction

By A. D. BUCKINGHAM† and J. A. POPLE‡
†Physical Chemistry Laboratory, University of Oxford
‡Department of Theoretical Chemistry, University of Cambridge

MS. received 30th April 1956

Abstract. A theory of the birefringence produced in a substance perpendicular to a strong magnetic field (the Cotton–Mouton effect) is presented. Allowance is made for electrical anisotropy induced directly by the magnetic field, an effect which can occur even for isotropic molecules. The effect of molecular interaction on measurements in dense gases and liquids is also considered.

---

§ 1. INTRODUCTION

MANY isotropic substances are weakly birefringent when light is passed through them in a direction at right angles to a strong magnetic field. The phenomenon is called the *Cotton–Mouton effect*, and it is the magnetic analogue of the Kerr electro-optical effect. The Cotton–Mouton effect should not be confused with the well-known Faraday effect, in which there is a rotation of the plane of polarization of a beam of light passing through a substance *parallel* to the lines of force of a magnetic field. The history and early experimental results of the Cotton–Mouton effect have been discussed by Beams (1932) and by Cotton (1932). Much experimental work has been carried out on diamagnetic liquids, and benzene has been studied as a vapour (König 1938); also the paramagnetic gases oxygen and nitric oxide have been examined and the results discussed (Bizette 1946).

The effect in diamagnetic substances has been interpreted by supposing that the molecules have anisotropic electrical and magnetic polarizabilities, so that the presence of a strong uniform magnetic field causes a partial lining-up of the molecules, the medium therefore becoming doubly refracting, with its optic axis parallel to the magnetic field. There is thus a close analogy to the Kerr effect, in which a strong electric field favours certain molecular configurations thereby causing double refraction. The accepted explanation of the Kerr effect was that of Langevin (1910) and Born (1918), who considered the orienting influence of the electric field on the anisotropic polarizabilities and permanent dipole moments of the molecules. Recently, Buckingham and Pople (1955) interpreted the observed non-zero Kerr constants of the inert gases in terms of an electrical 'hyperpolarizability' representing the dependence of the polarizability of a molecule on the field strength.

The Langevin theory of the electric double refraction has been used to interpret the Cotton–Mouton effect in diamagnetic molecules, and values for the principal diamagnetic susceptibilities of molecules deduced from the experimental results (see Beams 1932). However, the Langevin theory is valid only for perfect gases, so that there is some uncertainty in the deductions. Attempts to estimate the importance of molecular interactions have been made by Raman and Krishnan (1927), Piekara (1938, 1950) and Snellman (1949) but none of these is entirely satisfactory.

In the present paper, it is shown how the general theory of molecular polarizability in the presence of a strong magnetic field can throw light on the anisotropy induced by the magnetic field. If an atom is placed in a strong magnetic field H, then the electric dipole moments induced by a weak field E are given by expansions of the form

$$\mu = (\alpha + \eta^{\parallel} H^2 + \ldots)\, E, \qquad \ldots\ldots(1.1)$$

$$\mu = (\alpha + \eta^{\perp} H^2 + \ldots)\, E, \qquad \ldots\ldots(1.2)$$

according to whether E is parallel or perpendicular to H. α is the normal electrical polarizability and $\eta^{\parallel}$ and $\eta^{\perp}$ are coefficients representing the modification of the usual law due to the presence of the magnetic field. This means that there will be a Cotton–Mouton effect even for a monatomic substance. The Langevin theory of the Cotton–Mouton effect is modified so as to include direct polarizability effects of this type.

§ 2. THE POLARIZATION OF A MOLECULE IN A STRONG ELECTROMAGNETIC FIELD

Although the Cotton–Mouton effect is dependent upon the electrical molecular polarizability in the alternating electric field of a light wave, we shall suppose that the problem is equivalent to that of the static electric polarizability of a molecule in a magnetic field. We shall therefore neglect any dispersion effects; if the frequency is not close to a natural frequency of the molecule, this should be a reasonable approximation. We are particularly concerned with the change in electric polarizability caused by the magnetic field, and shall therefore start by formulating a general method of describing this compound polarizability.

Suppose we are dealing with a diamagnetic molecule in a closed-shell ground state (so that there is no Zeeman splitting in the magnetic field). Then, if we treat the translational and rotational degrees of freedom in a classical manner, there will be an energy $u(\tau, \mathbf{E}, \mathbf{H})$ corresponding to a molecule whose configuration (position and orientation) is represented by the symbol τ, in an electric field $\mathbf{E}$ and a magnetic field $\mathbf{H}$. This can be expanded in a power series in E and H

$$u(\tau, \mathbf{E}, \mathbf{H}) = u^{(0)} - \mu_\alpha{}^{(0)} E_\alpha - \tfrac{1}{2}\alpha_{\alpha\beta} E_\alpha E_\beta - \tfrac{1}{2}\chi_{\alpha\beta} H_\alpha H_\beta$$

$$- \xi_{\alpha:\beta\gamma} E_\alpha H_\beta H_\gamma - \tfrac{1}{2}\eta_{\alpha\beta:\gamma\delta} E_\alpha E_\beta H_\gamma H_\delta + \text{terms } O(E^3, H^3), \qquad \ldots\ldots(2.1)$$

where $\mu_\alpha{}^{(0)}$, $\alpha_{\alpha\beta}$, $\chi_{\alpha\beta}$, $\xi_{\alpha:\beta\gamma}$ and $\eta_{\alpha\beta:\gamma\delta}$ are tensors. $\mu_\alpha{}^{(0)}$ is the permanent electric dipole moment and $\alpha_{\alpha\beta}$ and $\chi_{\alpha\beta}$ are the static electric and magnetic susceptibility tensors. $\xi_{\alpha:\beta\gamma}$ and $\eta_{\alpha\beta:\gamma\delta}$ are higher order tensors representing the effect of a magnetic field on the electrical properties. $\xi_{\alpha:\beta\gamma}$ will be symmetrical in the suffices β and γ and $\eta_{\alpha\beta:\gamma\delta}$ symmetrical in the separate pairs (α, β) and (γ, δ).

The electric dipole moment μ_α (in the presence of $\mathbf{E}$ and $\mathbf{H}$) may be found from the relation

$$du = -\mu_\alpha \, dE_\alpha \qquad \ldots\ldots(2.2)$$

so that

$$\mu_\alpha(\tau, \mathbf{E}, \mathbf{H}) = \mu_\alpha{}^{(0)} + \alpha_{\alpha\beta} E_\beta + \xi_{\alpha:\beta\gamma} H_\beta H_\gamma + \eta_{\alpha\beta:\gamma\delta} E_\beta H_\gamma H_\delta + \ldots \qquad \ldots\ldots(2.3)$$

The tensor $\xi_{\alpha:\beta\gamma}$ clearly measures the change in electric dipole moment induced by the magnetic field (an effect proportional to H^2). We are interested in the electric polarizability in the magnetic field. This is obtained by further differentiation with respect to E and is

$$\Pi_{\alpha\beta}=\alpha_{\alpha\beta}+\eta_{\alpha\beta:\gamma\delta}H_\gamma H_\delta+\ldots \qquad \ldots\ldots(2.4)$$

As in the corresponding theory of the Kerr effect, $\Pi_{\alpha\beta}$ will be referred to as the *differential polarizability*; the tensor $\eta_{\alpha\beta:\gamma\delta}$ measures its dependence on H. It is analogous to the hyperpolarizability tensor $\gamma_{\alpha\beta\gamma\delta}$ introduced into the corresponding theory of the Kerr effect to allow for the dependence of electric polarizability on electric field strength (Buckingham and Pople 1955).

For symmetrical molecules there are relations between the components of these tensors. In particular, for a system with spherical symmetry $\xi_{\alpha:\beta\gamma}$ vanishes and $\eta_{\alpha\beta:\gamma\delta}$ is an isotropic tensor of order four. Since it is symmetrical in the pairs of suffices (α, β) and (γ, δ) it can be written in the form

$$\eta_{\alpha\beta:\gamma\delta}=\lambda\delta_{\alpha\beta}\delta_{\gamma\delta}+\mu(\delta_{\alpha\gamma}\delta_{\beta\delta}+\delta_{\alpha\delta}\delta_{\beta\gamma}). \qquad \ldots\ldots(2.5)$$

Consequently all components vanish except those of the types

$$\begin{aligned}\eta_{11:11}&=\eta^{\parallel}=\lambda+2\mu\\ \eta_{11:22}&=\eta^{\perp}=\lambda\\ \eta_{12:12}&=\tfrac{1}{2}(\eta^{\parallel}-\eta^{\perp})=\mu.\end{aligned} \qquad \ldots\ldots(2.6)$$

$\eta^{\parallel}$ is proportional to the change in polarizability parallel to the magnetic field and $\eta^{\perp}$ is the corresponding constant for the perpendicular direction.

§ 3. THEORETICAL EXPRESSION FOR THE MOLECULAR COTTON-MOUTON CONSTANT

The quantity measured in the Cotton–Mouton experiment is the difference between the refractive indices of a substance in directions parallel and perpendicular to a strong magnetic field H_α. Let $e_\alpha^{\parallel}$ and $e_\alpha^{\perp}$ be unit vectors in these two directions. If the frequency of the light waves is large compared with rotational frequencies, we can obtain an expression for this difference by evaluating

$$\Pi(\tau,\mathbf{H})=\Pi_{\alpha\beta}(e_\alpha^{\parallel}e_\beta^{\parallel}-e_\alpha^{\perp}e_\beta^{\perp}) \qquad \ldots\ldots(3.1)$$

for each configuration τ and then averaging over configurations with a Boltzmann type weighting factor appropriate to the molecule in the static field H_α only. The difference between the refractive indices is given by

$$n_{\parallel}-n_{\perp}=2\pi(N/V)\overline{\Pi} \qquad \ldots\ldots(3.2)$$

where N/V is the number of molecules in unit volume, and where

$$\overline{\Pi}=\frac{\int\Pi(\tau,\mathbf{H})\exp\{-u(\tau,\mathbf{H})/kT\}\,d\tau}{\int\exp\{-u(\tau,\mathbf{H})/kT\}\,d\tau}. \qquad \ldots\ldots(3.3)$$

We now proceed to expand $\overline{\Pi}$ as a power series in H, noting that it depends on H both through $\Pi(\tau,\mathbf{H})$ and through the energy u. It is convenient to use the notation $\langle X\rangle$ to denote the average value of a quantity $X(\tau,\mathbf{H})$ in the assembly with $H=0$, that is,

$$\langle X\rangle=\frac{\int X(\tau,0)\exp\{-u^{(0)}/kT\}\,d\tau}{\int\exp\{-u^{(0)}/kT\}\,d\tau}. \qquad \ldots\ldots(3.4)$$

It is seen immediately that $\overline{\Pi}$ is zero when $H=0$, that is the medium is isotropic in the absence of the magnetic field. Also

$$\left(\frac{\partial\overline{\Pi}}{\partial H}\right)_{H=0}=\left\langle\frac{\partial\Pi}{\partial H}\right\rangle-\frac{1}{kT}\left\langle\Pi\frac{\partial u}{\partial H}\right\rangle=0, \qquad \ldots\ldots(3.5)$$

since both $(\partial\Pi/\partial H)_{H=0}$ and $(\partial u/\partial H)_{H=0}$ are zero (from (3.1), (2.4) and (2.1)). Hence if we expand $\overline{\Pi}$ in powers of H, the leading term is in H^2, the coefficient being

$$\tfrac{1}{2}\left(\frac{\partial^2\overline{\Pi}}{\partial H^2}\right)_{H=0}=\tfrac{1}{2}\left\langle\frac{\partial^2\Pi}{\partial H^2}\right\rangle-\frac{1}{2kT}\left\langle 2\frac{\partial\Pi}{\partial H}\frac{\partial u}{\partial H}+\Pi\frac{\partial^2 u}{\partial H^2}\right\rangle+\frac{1}{2k^2T^2}\left\langle\Pi\left(\frac{\partial u}{\partial H}\right)^2\right\rangle. \qquad \ldots\ldots(3.6)$$

Now

$$\left(\frac{\partial^2\Pi}{\partial H^2}\right)_{H=0}=\left(\frac{\partial^2\Pi_{\alpha\beta}}{\partial H^2}\right)_{H=0}(e_\alpha{}^{\parallel}e_\beta{}^{\parallel}-e_\alpha{}^{\perp}e_\beta{}^{\perp})$$

$$=2\eta_{\alpha\beta:\gamma\delta}(e_\alpha{}^{\parallel}e_\beta{}^{\parallel}-e_\alpha{}^{\perp}e_\beta{}^{\perp})e_\gamma{}^{\parallel}e_\delta{}^{\parallel} \qquad \ldots\ldots(3.7)$$

and

$$\left(\frac{\partial^2 u}{\partial H^2}\right)_{H=0}=-\chi_{\alpha\beta}e_\alpha{}^{\parallel}e_\beta{}^{\parallel}. \qquad \ldots\ldots(3.8)$$

We now use $(\Pi)_{H=0}=\alpha_{\alpha\beta}(e_\alpha{}^{\parallel}e_\beta{}^{\parallel}-e_\alpha{}^{\perp}e_\beta{}^{\perp})$ and carry out the final averaging over directions as in the theory of the Kerr constant (Buckingham and Pople 1955).

Analogously to the molecular Kerr constant, we shall define a *molecular Cotton–Mouton* constant ${}_{\mathrm{m}}C$ by

$${}_{\mathrm{m}}C=\lim_{H\to 0}\frac{2(n_{\parallel}-n_{\perp})V_{\mathrm{m}}}{27H^2}=\frac{2\pi N}{27}\left(\frac{\partial^2\overline{\Pi}}{\partial H^2}\right)_{H=0} \qquad \ldots\ldots(3.9)$$

where V_{m} is the molar volume and N Avogadro's number. ${}_{\mathrm{m}}C$ is one-ninth of a similar quantity defined by König (1938) and six times that introduced by Grodde (1938). From (3.9) and (3.6) we obtain

$${}_{\mathrm{m}}C=\frac{4\pi N}{135}\left\{\eta_{\alpha\beta:\alpha\beta}-\tfrac{1}{3}\eta_{\alpha\alpha:\beta\beta}+\frac{1}{2kT}\Big[\alpha_{\alpha\beta}\chi_{\alpha\beta}-3\alpha\chi\Big]\right\} \qquad \ldots\ldots(3.10)$$

where $\alpha=\frac{1}{3}\alpha_{\alpha\alpha}$ and $\chi=\frac{1}{3}\chi_{\alpha\alpha}$ are the mean electric and magnetic susceptibilities. The temperature-dependent part of this formula is the established Langevin result. It is non-zero only if both $\alpha_{\alpha\beta}$ and $\chi_{\alpha\beta}$ are anisotropic. The other terms represent the molecular anisotropy induced directly by the magnetic field. For spherical molecules these are the only terms and (3.10) reduces to

$${}_{\mathrm{m}}C=\frac{4\pi N}{27}(\eta^{\parallel}-\eta^{\perp}). \qquad \ldots\ldots(3.11)$$

§ 4. THE INFLUENCE OF DENSITY

In a dense medium ${}_{\mathrm{m}}C$ is defined for a system in which H is uniform by the equation

$${}_{\mathrm{m}}C=\frac{2nV_{\mathrm{m}}}{3(n^2+2)^2}\lim_{H\to 0}\frac{n_{\parallel}-n_{\perp}}{H^2} \qquad \ldots\ldots(4.1)$$

where n is the refractive index of the medium. Then, just as in the theory of the molecular Kerr constant in a dense medium (Buckingham 1955),

$${}_{\mathrm{m}}C=\frac{2\pi N}{27}\left(\frac{\partial^2\overline{\Pi^{(1)}}}{\partial H^2}\right)_{H=0} \qquad \ldots\ldots(4.2)$$

where $\overline{\Pi^{(1)}}$ is the mean value of $\Pi(\tau, \mathbf{H})$ for molecule 1; the symbol τ now represents the configuration of all the molecules of the system.

If we now expand $\overline{\Pi^{(1)}}$ in powers of $\chi_{\alpha\beta}$ and neglect powers higher than the first, then by a method similar to that employed elsewhere in the Kerr effect problem (Buckingham 1955) we obtain

$$_{\mathrm{m}}C = \frac{4\pi N}{135}\left\{\eta^{(1)}{}_{\alpha\beta:\alpha\beta} - \tfrac{1}{3}\eta^{(1)}{}_{\alpha\alpha:\beta\beta} + \frac{1}{2kT}\sum_i \left[\alpha_{\alpha\beta}{}^{(1)}\chi_{\alpha\beta}{}^{(i)} - 3\alpha^{(1)}\chi^{(i)}\right]\right\} \qquad \ldots\ldots(4.3)$$

where the summation extends over all the molecules of the specimen. If the molecules are axially symmetrical, then the two independent $\alpha_{\alpha\beta}$ and $\chi_{\alpha\beta}$ tensors can be written in terms of α and χ and of the anisotropies

$$\kappa_\alpha = \frac{\alpha_3 - \alpha_1}{3\alpha}, \quad \kappa_\chi = \frac{\chi_3 - \chi_1}{3\chi} \qquad \ldots\ldots(4.4)$$

where the 3-direction corresponds to the axis of symmetry. In terms of these quantities (4.3) becomes

$$_{\mathrm{m}}C = \frac{4\pi N}{135}\left\{\eta_{\alpha\beta:\alpha\beta} - \tfrac{1}{3}\eta_{\alpha\alpha:\beta\beta} + \frac{3\alpha\chi\kappa_\alpha\kappa_\chi}{kT}\left[1 + \tfrac{3}{2}\int\!\!\int(\cos^2\theta - \tfrac{1}{3})n_2(\mathbf{r}, \omega)\, d\mathbf{r}\, d\omega\right]\right\} \qquad \ldots\ldots(4.5)$$

where θ is the angle between the directions of the axes of molecule 1 and the molecule whose configuration is $(\mathbf{r}, \omega)$ relative to molecule 1, and where $n_2(\mathbf{r}, \omega)\, d\mathbf{r}\, d\omega$ is the probability of there being a molecule in the volume element $d\mathbf{r}$ and orientational element $d\omega$. The integration extends over the whole of the sample. If the forces between the molecules are independent of relative orientation, then $\langle\cos^2\theta - \tfrac{1}{3}\rangle$ is zero and (4.5) reduces to

$$_{\mathrm{m}}C = \frac{4\pi N}{135}\left\{\eta_{\alpha\beta:\alpha\chi} - \tfrac{1}{3}\eta_{\alpha\alpha:\beta\beta} + \frac{3\alpha\chi\kappa_\alpha\kappa_\chi}{kT}\right\} \qquad \ldots\ldots(4.6)$$

which is equivalent to the perfect gas expression (3.10). Thus for diamagnetic materials we should expect $_{\mathrm{m}}C$ to depend on density only if the intermolecular forces are significantly non-central. If the forces are orientationally dependent, then it might be possible to estimate $\langle\cos^2\theta - \tfrac{1}{3}\rangle$ from independent observations on the Kerr effect, on the anisotropy induced by an intense beam of light (see Buckingham 1956) or on the depolarization of scattered light (see Benoit and Stockmayer 1956), for the deviations of all of these phenomena from their respective perfect gas values are approximately proportional to this factor.

§ 5. DISCUSSION

There is not a great deal of evidence about the magnitude of the direct polarization term in the theoretical expression for $_{\mathrm{m}}C$. There do not appear to be any measurements on monatomic substances for which the temperature-dependent term is zero. A theoretical calculation on atomic hydrogen can be used to get some idea of the order of magnitude for atoms. An exact calculation (neglecting the effect of electron spin) gives (Buckingham and Pople 1956)

$$\eta^{\parallel} = -\frac{159}{16}\frac{a^6}{mc^2}, \qquad \eta^{\perp} = -\frac{797}{48}\frac{a^6}{mc^2} \qquad \ldots\ldots(5.1)$$

where energies are measured in ergs, electric fields in e.s.u., and magnetic fields in oersteds; a is the Bohr radius and m the mass of the electron. Substitution in (3.11) gives ${}_{\mathrm{m}}C = 0{\cdot}05 \times 10^{-18}$. The value for one of the heavier inert gases may be about ten times larger (the ordinary diamagnetic susceptibility increases by this order of magnitude) but this is still considerably smaller than values for typical molecular substances. For example at 20°C liquid paraffin has ${}_{\mathrm{m}}C = -50 \times 10^{-18}$, and benzene ${}_{\mathrm{m}}C = 183 \times 10^{-18}$. However, it may be observable with present techniques.

There is some experimental evidence that the direct term may be important for conjugated substances which generally show a large Cotton–Mouton constant. Goldet (1935) and Grodde (1938) showed that the product ${}_{\mathrm{m}}CT$ was not independent of T for several liquids. Grodde's measurements of ${}_{\mathrm{m}}C$ and the molecular Kerr constant ${}_{\mathrm{m}}K$ for liquids indicate that the mean value $\langle \cos^2\theta - \frac{1}{3} \rangle$ in (4.5) is often small, especially when the molecules are non-polar. If this is neglected, the failure of ${}_{\mathrm{m}}C$ to be proportional to $1/T$ can be attributed to the temperature-independent term in (4.6). For 1 : 3 : 5 trimethylbenzene, for example, Grodde's data in the range 20–120°C can be represented by the equation

$${}_{\mathrm{m}}C = (-10 + 11500/T) \times 10^{-18}.$$

This would imply that the direct polarization term is negative for this substance.

ACKNOWLEDGMENT

One of us (A.D.B.) is indebted to the Commissioners for the Exhibition of 1851 for the award of a Senior Studentship.

REFERENCES

BEAMS, J. W., 1932, *Rev. Mod. Phys.*, **4**, 133.

BENOIT, H., and STOCKMAYER, W. H., 1956, *J. Phys. Radium*, **17**, 21.

BIZETTE, H., 1946, *Ann. Phys., Paris*, (12), **1**, 233.

BORN, M., 1918, *Ann. Phys., Lpz.*, **55**, 177.

BUCKINGHAM, A. D., 1955, *Proc. Phys. Soc.* A, **68**, 910; 1956, *Proc. Phys. Soc.* B, **69**, 344.

BUCKINGHAM, A. D., and POPLE, J. A., 1955, *Proc. Phys. Soc.* A, **68**, 905; 1956 *Proc. Camb. Phil. Soc.*, in the press.

COTTON, A., 1932, *Rapp. Cons. Phys. Solvay*, p. 418.

GOLDET, A., 1935, *C.R. Acad. Sci., Paris*, **200**, 654.

GRODDE, K.-H., 1938, *Phys. Z.*, **39**, 772.

KÖNIG, H., 1938, *Ann. Phys., Lpz.*, **31**, 289.

LANGEVIN, P., 1910, *Radium, Paris*, **7**, 249.

PIEKARA, A., 1938, *Z. Phys.*, **108**, 395; 1950, *Acta Phys. Polon.*, **10**, 37, 107.

RAMAN, C. V., and KRISHNAN, K. S., 1927, *Proc. Roy. Soc.* A, **117**, 1.

SNELLMAN, O., 1949, *Phil. Mag.*, **40**, 983.

Volume Generated Currents and Secondary Effects in Amorphous Selenium Films

By W. E. SPEAR

Physics Department, University College of Leicester

MS. received 28th March 1956

Abstract. Charge carriers are generated in amorphous Se films by electron bombardment. The films, prepared by evaporation, are between 3 and 9 μ thick and carry Al electrodes. The resulting current is studied in relation to the average depth of penetration of the bombarding beam. The experiments confirm the exceptionally large effective mobility of positive holes reported by Weimer. It is shown that secondary effects, which cause the entry of additional charge carriers from the electrodes normally play an important part. Saturation of the measured current with applied voltage is observed only with small bombarding currents when the generation of electron–hole pairs is confined to a narrow region near the top electrode. When the excitation extends throughout the specimen the use of a pulsed bombarding beam makes it possible to distinguish the true volume generated current from that due to secondary effects. The experiments lead to values between 18 and 25 ev for the average energy used in the production of an electron–hole pair. The mechanism whereby secondary carriers are drawn from the electrodes is briefly discussed.

---

§ 1. Introduction

Although the photoelectric properties of grey (hexagonal) Se have been studied for many years (e.g. Hughes and DuBridge 1932) little attention appears to have been given to the highly insulating amorphous form. Comparatively recently Weimer (1950) and Pensak (1950) drew attention to the remarkably long ranges of positive holes in this substance. This property makes amorphous Se particularly suitable for the study of the problem of saturation of volume generated currents with applied voltage. The absence of saturation at high current gains observed in experiments on amorphous and microcrystalline films (Ansbacher and Ehrenberg 1951) has been one of the difficulties in the interpretation of the results. This is connected with the fact first pointed out by Gudden and Pohl (e.g. 1925) that the fundamental primary process of carrier generation is normally accompanied by secondary effects, which cause the entry of additional (secondary) charge carriers from the electrodes.

In the experiments described in this paper charge carriers were generated by electron bombardment. This makes it possible to study the resulting current when the region of excitation extends to various depths within the film (Spear 1955, to be referred to as paper I). It was found that in the case of Se secondary effects often play a predominant part, so that the main problem consisted in isolating the volume generated component of the measured current. This was possible under the experimental conditions described in § 4.4 and 4.5. The saturation of the volume generated current with applied voltage was observed in both these experiments.

§ 2. EXPERIMENTAL ARRANGEMENT

The apparatus was essentially the same as that described in paper I. The bombarding electron beam, focused magnetically to a circular spot of 0·26 cm diameter, entered the specimen through the evaporated top electrode T. Energy losses in T were normally negligible. Figure 1 (*a*) shows the basic measuring circuit used for experiments with steady electron bombardment. The total current flowing in or out of the bottom electrode B was measured with a sensitive galvanometer. A simple d.c. amplifier with electrometer valve was used for the measurement of the small bombarding currents. The results with a pulsed primary beam were obtained with a bi-stable multivibrator unit connected to the control grid of the electron gun. The unit was kept at the potential of the gun (up to 50 kv) and the trigger pulses were applied by a high voltage coupling condenser.

§ 3. PREPARATION OF SPECIMENS

Specimens were prepared by evaporation from spectroscopically standardized grey Se shot.† Film thicknesses d ranged from 3 to 9 μ. Normally the Se was deposited on to a cleaved mica base carrying the evaporated aluminium bottom electrode. Most experiments were made using specimens with aluminium top electrodes, but some carrying gold and indium electrodes were investigated. A few 'transparent' specimens were prepared in connection with the determination of V_0 (§ 4.2). Mica cleaved to a thickness of about 0·3 μ was covered on both sides by thin evaporated aluminium. A base of this kind is surprisingly strong and practically transparent to electrons of energies considered in this paper.

§ 4. EXPERIMENTAL RESULTS

Altogether 19 specimens were investigated, chosen from ten different evaporations. The set of experimental results described in the following sections was obtained with a specimen of medium thickness (5·7 μ) and represents a typical example of the behaviour of amorphous Se films with aluminium electrodes.

4.1. *The Dark Current*

In absence of electron bombardment a dark current, i_d, flows, which in the case of Se films increases approximately as the sixth power of the applied voltage v. (i_d, v) curves of this kind have been observed for a number of amorphous and crystalline dielectrics (von Hippel 1938, Smith and Rose 1955). Weimer and Cope (1951) have given fairly conclusive evidence that positive holes can be drawn into the Se from a metal electrode or an evaporated Sb_2S_3 film in contact with the Se. No such effect was observed for electrons. It is, therefore, most likely that in the experiments described here i_d is predominantly carried by positive holes drawn from the positive electrode.

It was noted with earlier specimens that the dark current measured with a positive top electrode was 10–20 times larger than that with the same positive potential applied to the bottom electrode. This asymmetry was reduced substantially (see i_d curves in figure 2) by evaporating the top electrode at such a rate that the measured temperature at the surface remained below 45°c. It is likely that with faster evaporation local heating causes a partial crystallization to the far more conductive monoclinic or hexagonal forms. The above effect is connected with the top electrode and is only observed when this is at a positive potential. The observation is, therefore, consistent with the entry of holes from the top electrode.

---

† 'H.S.' Brand, supplied by Messrs. Johnson, Matthey & Co., Ltd.

After some initial bombardment fairly consistent values of i_d are obtained. The life of a normal specimen is about two hours under electron bombardment; then the dark currents starts to increase gradually. The same happens when fields too close to the breakdown value ($\simeq 4 \times 10^5$ v cm^{-1}) are applied.

4.2. *Electron Penetration Experiments*

In these experiments i_m, the change in current through the specimen produced by the electron bombardment is observed as a function of V_p, the energy of the primary electrons. i_m is obtained from $i_m = i_B - i_d - i_a$, where i_B is the total current flowing in or out of electrode B, and i_a is the absorbed part of the primary current i_p. The sign of the applied potential v (the bias) is defined as that of electrode B. In the following a subscript is frequently used to denote the sign of v, e.g. $(i_m)_-$, $(h_2)_+$.

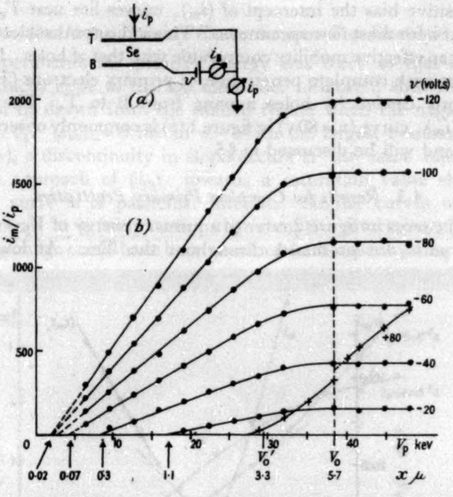


Figure 1. (*a*) Basic measuring circuit. (*b*) Graph of i_m/i_p as a function of V_p for different values of v. $i_p = 1{\cdot}4 \times 10^{-8}$ A. Numbers below V_p axis are calculated values of the average depth of penetration of the bombarding electrons.

At zero bias, the Se films behaved very much like the dielectrics investigated in paper I. At $V_p = V_0'$ an electron current begins to flow from B to earth. This occurs before V_0, the energy at which primary electrons begin to pass through the film. The relation $V_0^2 = bm$ (m = mass per unit area of film) is found to apply with $b = 6{\cdot}1 \times 10^{11}$ v^2 g^{-1} cm^2. Some calculated values of x, the average depth of penetration of the primary beam, are given in figure 1 (*b*). At $V_p = V_0'$,

$x' = 0{\cdot}57d$. Both b and x'/d fall within the range of values obtained for other dielectrics.

An interesting feature shown in figure 1 (*b*) is the shift of the intercept of the $(i_m)_-$ curves as the negative bias is increased. This is a special property of Se films. For all dielectrics investigated so far both $(i_m)_-$ and $(i_m)_+$ curves cut the V_p axis at V_0', independently of magnitude and polarity of v. But in Se with, for example $v = -40$ v, a current begins to flow through the specimen when the primary beam produces electron–hole pairs down to a depth of about $0{\cdot}3\,\mu$. The sign of the bias shows that the observed current is carried by positive holes. At $v = -100$ v holes created practically at the top electrode will reach B and no further shift of the intercept is found at $v = -120$ v and with higher bias.

$(i_m)_-$ curves rise almost linearly with V_p; as V_0 is approached the current tends towards a constant value. The current gains at room temperature are several hundred times larger than for arsenic or antimony sulphide films at comparable values of v/d. These and the following experimental results are consistent with the idea of an exceptionally large effective mobility of positive holes in amorphous Se. ('Effective' is used to imply that the mobility includes the effect of trapping).

With a positive bias the intercept of $(i_m)_+$ curves lies near V_0' and is fairly independent of v for most new specimens. This indicates that electrons certainly do not possess an effective mobility comparable with that of holes. It is, therefore, most likely that with complete penetration of primary electrons ($V_p > V_0$), $(i_m)_+$ is predominantly carried by holes moving from B to T. The discontinuity shown by the $(i_m)_+$ curve ($v = 80$ v) in figure 1 (*b*) is commonly observed at positive and zero bias and will be discussed in § 5.

4.3. *Results for Complete Primary Penetration*

The (i_m, v) curves in figure 2 refer to a primary energy of $V_p = 43$ kev, so that electron–hole pairs are produced throughout the film. At low bombarding

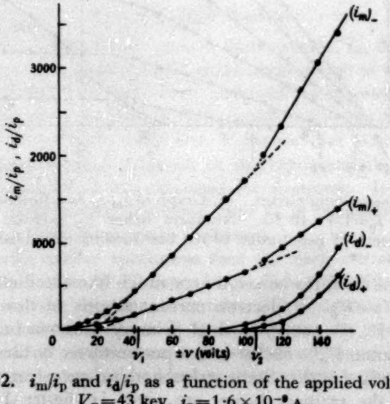


Figure 2. i_m/i_p and i_d/i_p as a function of the applied voltage v. $V_p = 43$ kev, $i_p = 1{\cdot}6 \times 10^{-8}$ A.

currents ($i_p \simeq 10^{-9}$ A) most of the observed curves could be divided into three distinct parts: (i) an initial curved part, (ii) a practically linear part between v_1 and v_2, and (iii) a rapidly increasing part separated from (ii) by a change of slope at v_2. At higher bombarding currents the division between these ranges becomes less distinct; at $i_p \simeq 10^{-8}$ A it is no longer apparent.

$(i_m)_-$ and $(i_m)_+$ curves are very similar; both show the characteristic changes at v_1 and v_2. The value of v_1 is normally the same for positive and negative bias and is roughly proportional to d. The position of v_2 is connected with the behaviour of the dark current as indicated by the following observations: (i) For a specimen with approximately symmetric i_d the values of $(v_2)_+$ and $(v_2)_-$ are in reasonable agreement. (ii) If for any reason the dark current in one direction, say $(i_d)_-$, is permanently increased (e.g. by applying too large a potential), then a shift of $(v_2)_-$ towards v_1 is observed. The position of $(v_2)_+$ is not normally affected. (iii) A few specimens showed an exceptionally low value of $(i_d)_+$. The $(i_m)_+$ curve did not initially possess the discontinuity at v_2. This developed only when after some electron bombardment the dark current had increased. These observations show fairly conclusively that the discontinuity at v_2 is connected with the entry of additional carriers from the electrodes.

4.4. *Results for Primary Excitation near Top Electrode*

In these experiments the primary energy was 6 kev so that electron–hole pairs were produced close to the top electrode. Figure 3 shows that a steady current can just be drawn from the excited region when the negative bias has a value close to v_1. $(i_m)_+ = 0$ for all v. With the largest bombarding current ($i_p = 1{\cdot}6 \times 10^{-8}$ A), a discontinuity in slope occurs at the same value of v_2 as in figure 2. The approach of $(i_m)_-$ towards a saturation value shown by the $i_p = 1{\cdot}6 \times 10^{-9}$ A curve is of particular interest. Similar curves were obtained for most specimens; it is fairly certain that complete saturation of $(i_m)_-$ within the given voltage range could have been obtained in all cases by a further decrease

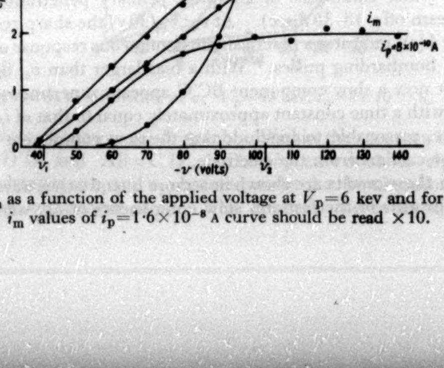


Figure 3. i_m as a function of the applied voltage at $V_p = 6$ kev and for different i_p. Note: i_m values of $i_p = 1{\cdot}6 \times 10^{-8}$ A curve should be read ×10.

in i_p. But this is limited by the fact that, at higher v, $i_d \gg i_m$, which makes it difficult to measure i_m with sufficient accuracy. In the case of a few specimens with comparatively low and steady dark current, it has been possible to observe the complete saturation of i_m with v. This is shown by the $i_p = 8 \times 10^{-10}$ A curve in figure 3. Although at $v = -140$ v i_m amounts to only 7% of i_d, it was possible to obtain reasonably consistent values by measuring i_B and i_d in quick succession. The points plotted in figure 3 are the mean values of four such measurements.

As the energy of the incident electrons W_p is completely absorbed in a thin layer, all the charge carriers produced in this region contribute equally to the measured saturation current. Thus the average energy required for the production of one electron–hole pair is $\epsilon = \mathbf{e}W_p/(i_m)$. The value of ϵ, obtained from specimens which gave a definite saturation value, lies between 18 and 23 ev which is of the same order as that obtained by McKay (1950) for diamond under comparable experimental conditions.

With steady electron bombardment saturation currents have been observed only if a small number of carriers are generated in a region close to the top electrode. The above measurements are, therefore, dependent on whether under these conditions a sufficient number of charge carriers can be drawn across the complete specimen at fields substantially below breakdown. As idicated in § 4.2 this has only been found possible with Se films and not with other amorphous or microcrystalline dielectrics. Soft x-rays generated at the top electrode might affect the value of the saturation current (Firmin and Oatley 1955). To test this an Al foil of sufficient thickness to absorb all incident electrons was mounted above T. Even with i_p as large as 10^{-7} A no measurable effect could be detected.

4.5. *Pulse Measurements*

Figures 4 (*a*) and 4 (*b*) show typical pulse shapes when in absence of electron bombardment a pulsating bias is applied across the specimen. (Bias on: from 0–18 msec; bias off: from 18–36 msec.) At a pulse height of $v = 40$ v no measurable current flows between 0·3 and 18 msec. The initial rise of i_d could just be resolved from the displacement current peak if 0·5 msec pulses were used. (Time constant of circuit $\simeq 1\,\mu$sec.) However, with 100 v pulses figure 4 (*b*) shows the comparatively slow growth of i_d during the on-period. The pulse height at 18 msec is practically equal to the i_d value obtained with the same steady bias.

Figure 4 (*c*)–(*e*) illustrates results with a pulsating bombarding beam and a steady bias under conditions of complete primary penetration. (Beam on: 0–18 msec; beam off: 18–36 msec). At $v < v_1$ (30 v) the sharp rectangular pulse ABCD (figure 4 (*c*)) indicates a practically instantaneous response of the generated current to the bombarding pulses. With a bias larger than v_1, figures 4 (*d*) and 4 (*e*) show that now a slow component BC′C appears superimposed on the rectangular pulse with a time constant approximately equal to that of i_d in figure 4 (*b*). It is, therefore, reasonable to conclude that the slow component represents the entry of charge carriers from the electrode.

In figure 5 these results are shown in a more quantitative way. $h_1 = $ AB and $h_2 = $ C′C, the heights of the fast and slow components of i_m, were measured on

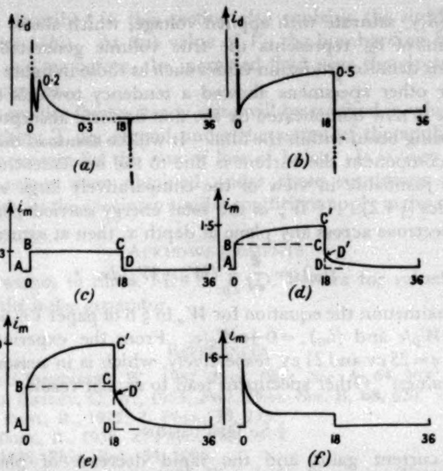


Figure 4. Observed pulse shapes. Time scale in msec, current values in μA. (*a*)–(*b*): Pulsed bias, no electron bombardment. (*a*) $v=40$ v, (*b*) $v=100$ v. (*c*)–(*e*): Pulsed electron beam, steady bias. $V_p=43$ kev, $\bar{i}_p=8\times10^{-10}$ A. (*c*) $v=30$ v, (*d*) $v=70$ v, (*e*) $v=100$ v. (*f*): Pulsed electron beam, steady bias. $V_p=10$ kev, $\bar{i}_p=4\times10^{-9}$ A, $v=45$ v.

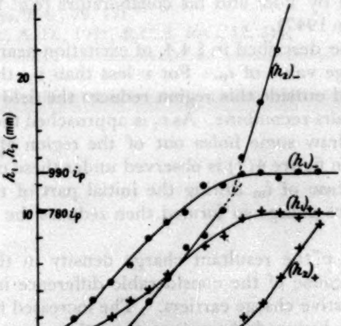


Figure 5. h_1 and h_2 as a function of v. $V_p=43$ kev, $\bar{i}_p=8\times10^{-10}$ A.

the screen of the cathode-ray oscillograph. Within the accuracy of the measurements C′C = D′D and AB = C′D′. The h_2 curves show clearly that the characteristic changes at v_1 and v_2 observed under d.c. bombardment are entirely due to the slow component of i_m. This confirms the conclusions reached in § 4.3.

$(h_1)_+$ and $(h_1)_-$ saturate with applied voltage, which shows conclusively that this component of i_m represents the true volume generated current. With seven specimens definite saturation values such as those in figure 5 were obtained; the curves for other specimens showed a tendency towards saturation. The calculation of ϵ is now complicated by the non-uniform absorption of the energy of the bombarding beam within the film. It will be assumed that during the fast rise of the h_1 component the current is due to the movement of carriers of one sign. This is justifiable in view of the comparatively large effective mobility of positive holes (§ 4.2). If W_x is the total energy carried per second by the bombarding electrons across any plane at depth x, then at saturation

$$(i_m)_- = \frac{e}{\epsilon d}\int_0^d (W_p - W_x)dx.$$

As a first approximation the equation for W_x in § 6 of paper I is used. This leads to $(i_m)_- = 0{\cdot}6eW_p/\epsilon$ and $(i_m)_+ = 0{\cdot}4eW_p/\epsilon$. From the experimental values of i_m in figure 5, $\epsilon = 25$ ev and 21 ev respectively, which is in reasonable agreement with the d.c. values. Other specimens lead to similar results.

§ 5. DISCUSSION

The large current gains and the rapid decrease of photocurrent with temperature (Weimer and Cope 1951) suggest that the large effective mobility of holes is connected with a comparatively low density of shallow traps. It is probable that the much smaller effective mobility of electrons is caused by a larger density or greater depth of electron traps. Based on this difference in effective mobility the observed secondary effects may be explained by a mechanism similar to that used by Pohl and his collaborators (e.g. Hilsch and Pohl 1938, Pohl and Stöckmann 1947).

Consider the case, described in § 4.4, of excitation near the top electrode with a comparatively large value of i_p. For v less than v_1, the space charge set up by the holes trapped outside this region reduces the field inside to such a value that electron–hole pairs recombine. As v_1 is approached the applied field becomes strong enough to draw some holes out of the region of recombination. The pulse shape shown in figure 4 (f) is observed under these conditions. It is likely that the rapid decrease of i_m during the initial part of the pulse is caused by trapping. The space charge so formed then reduces the current to the smaller constant value.

With increasing v the resultant charge density in the excited region will become negative because of the considerable difference in effective mobility of the positive and negative charge carriers. The increased field near the electrode will now draw in a larger dark current than that measured without electron bombardment. The field will grow until, in the steady state, the number of holes leaving the excited region per second is compensated by the rate at which electrons reach the electrode and the rate of recombination of electrons with the additional holes from the electrode. It is significant in this connection that the magnitude of v_2 is generally equal to that value of v at which, according to figure 1 (b), holes can just be drawn across the complete specimen. This suggests that for $v > v_2$ an increasing proportion of holes entering at T escapes recombination in the excited region. This should increase the field near T and lead to the larger influx of charge carriers from the electrode observed at higher bias voltages.

The mechanism described in this section also explains the sudden increase of $(i_m)_+$ at V_0 (figure 1 (*b*)). At this value of V_p the bombarding beam generates pairs close to B and, as before, the increased field near the electrode will draw in a larger dark current.

If i_p is made smaller then a steady state will be reached in which the number of electrons reaching T per second under the action of the applied field is just sufficient to compensate for the rate of loss of holes. It is very likely that the d.c. saturation values were obtained under these conditions and that with excitation throughout the specimen similar conditions apply in the range $0<v<v_1$.

ACKNOWLEDGMENTS

The author wishes to thank Mr. N. F. D. Pereira for valuable help in the construction of the pulse generator.

REFERENCES

ANSBACHER, F., and EHRENBERG, W., 1951, *Proc. Phys. Soc.* A, **64**, 362.
FIRMIN, J. C., and OATLEY, C. W., 1955, *Proc. Phys. Soc.* B, **68**, 620.
GUDDEN, B., and POHL, R., 1925, *Z. Phys.*, **35**, 243.
HILSCH, R., and POHL, R., 1938, *Z. Phys.*, **108**, 55.
VON HIPPEL, A., 1938, *Phys. Rev.*, **54**, 1096.
HUGHES, A. L., and DUBRIDGE, L. A., 1932, *Photoelectric Phenomena* (New York : McGraw-Hill), p. 330.
MCKAY, K. G., 1950, *Phys. Rev.*, **77**, 816.
PENSAK, L., 1950, *Phys. Rev.*, **79**, 171.
POHL, R., and STÖCKMANN, F., 1947, *Ann. Phys., Lpz.*, **6**, 275.
SMITH, R. W., and ROSE, A., 1955, *Phys. Rev.*, **97**, 1531.
SPEAR, W. E., 1955, *Proc. Phys. Soc.* B, **68**, 991.
WEIMER, P. K., 1950, *Phys. Rev.*, **79**, 171.
WEIMER, P. K., and COPE, A. D., 1951, *R.C.A. Rev.*, **12**, 314.

Bombardment of Various Elements by Hg^+ and A^+ Ions

By J. F. STRACHAN and N. L. HARRIS
Research Laboratories of The General Electric Company Limited, Wembley, Middx.

MS. received 9*th March* 1956 *and in revised form* 28*th May* 1956

Abstract. The sputtering, in low pressure discharge tubes, of carbon by Hg^+ and A^+ ions at energies from 10 ev to 20×10^3 ev was investigated by direct weight loss measurements. The commonly quoted linear variation of sputtering rate in grammes per ampere-hour held up to 2 to 3 kv, where it was replaced by a variation proportional to $(\text{voltage})^{1/2}$. Sputtering occurred at 30 volts, a value well below the usual 'critical voltages' in the literature. X-rays were generated above 8 kv by the action of secondary electrons emitted from the targets. The emission from spherical targets was determined from the dimensions of the positive ion sheaths, and was initially proportional to voltage, after which the rate of increase diminished rapidly.

A brief survey of the sputtering of beryllium, aluminium, silicon, iron, nickel, copper, molybdenum, tungsten, lead and uranium indicated similar variations with voltage, and showed that the sputtering rate was roughly proportional to the $(\text{atomic weight of the target})^{1\cdot 3}$ and that it increased with the heavier ion.

These empirical relations are useful for practical assessments, but not for theoretical interpretation, unless the secondary electron emission is distinguished. This was done for carbon and iron and the sputtering rate in atoms per ion s determined. The empirical relation of the form $s = \lambda \ln(V/V_0)$ was found, where V_0 is thought to be related to the 'penetration' of the ion into the target, and has a value of a few hundred volts. The ratio of the binding energy of the atom to the energy required to sputter it, for carbon and iron atoms released by Hg^+ and A^+ ions, had a maximum of 0·5% at 2000 volts for carbon and of 2·5% at 1000 volts for iron.

---

§ 1. Introduction

Sputtering was first observed about 1852, and has been investigated since then by many workers. Despite this lengthy study, it is still little understood, mainly because investigations were usually carried out with electrical discharge tubes, which are inherently associated with extremely complex physical processes.

The early investigations reported in the literature were reviewed by Compton and Langmuir (1930) who give an empirical relationship for the variation of sputtering rate with voltage of the type

$$w = W/it = \lambda(V - V_0) \qquad \ldots\ldots(1)$$

where w = mass of material sputtered in grammes per ampere-hour, W = total mass sputtered in grammes, i = specimen current in amperes, t = time in hours, λ = constant characteristic of gas, V = numerical value of negative voltage of specimen, V_0 = constant characteristic of sputtered surface, varying from 350 to 550 volts termed 'critical voltage'.

This expression resulted mainly from investigations of cathode sputtering in two-electrode tubes, where the cathode fall V was varied indirectly by adjustments of the gas pressure and tube current. Further, the gas pressures employed seem sufficiently high for collisions, charge transfer and back sputtering to have occurred. Thus the critical voltages given are likely to exceed the true threshold.

Improved experiments by Campbell (1923) at these laboratories and by Kingdon and Langmuir (1923) on the bombardment of tungsten and thoriated tungsten filaments respectively, in general confirmed this linear relationship but gave, by extrapolation of experimental results, critical voltages ranging from 30 to 130 for bombardment by mercury and argon ions. However, in both experiments the sputtering rates were determined indirectly and it is not certain that the observed changes were entirely due to sputtering. On the other hand Oliphant (1928) concluded from his results that critical voltage effects were characteristic of an adsorbed gas layer on the target, and that with thorough outgassing, only surface erosion was produced even up to several thousands of volts, without any loss of material by sputtering. He found no voltage below which this surface erosion did not occur.

Investigations by Meyer and Güntherschulze (1931) on the sputtering of tungsten, aluminium, molybdenum, nickel and iron in mercury vapour at voltages up to 4 kv showed approximately linear variations, while with tantalum up to 10 kv a pronounced departure from the linear law occurred at the higher voltages. Recently Güntherschulze (1955) has reconsidered some of his earlier experimental work, and has shown that for bombardment by mercury ions $w=\lambda(V^{1/2}-V_0^{1/2})$ where λ is a constant characteristic of the specimen. Further, he finds that for aluminium, silicon, manganese, iron, zirconium, molybdenum, tantalum and tungsten the value of V_0 is 190 volts, while for nickel it is 107 volts.

Many of the experimental problems and difficulties in interpretation inherent in ion bombardment of the cathode can be avoided by introducing a third electrode and bombarding it with ions drawn from the main discharge to cause the sputtering. This is the principle of the method employed in these investigations, and it had the additional merits that the target could also be used as a probe to investigate both its secondary electron emission and the conditions in the discharge. The presence of secondary electrons was revealed by the detection of x-rays at the higher bombarding voltages, and has been almost completely ignored in the literature of sputtering. Moreover, the discharge variables could be adjusted independently of the target so that the most stable conditions were attained for any given test. This method, which was well suited for its prime purpose, an investigation into the rate of loss of weight of various materials by sputtering over a wide range of voltages in mercury vapour and argon, in other words the macroscopic phenomenon, was not so well suited, however, for a fundamental study of the individual processes involved. On the other hand, high vacuum techniques to eliminate the complexities associated with discharge tubes would have raised the difficult practical problem of achieving an ion beam of sufficient density.

The relationship between sputtering and the energy of the bombarding ions was determined at low gas pressures, consistent with the maintenance of the discharge. The necessary condition is that the dimensions of the positive ion sheath in front of the target should be considerably less than the mean free path so that the probability of loss of energy by collisions with the gas atoms can be ignored. Under this condition sputtering is independent of pressure. At higher gas pressures, not only would the ions lose energy by collisions, but the probability of transfer of charge from the ions to atoms of the same species would be relatively high. Moreover, further complications due to 'back sputtering' could arise, especially if the gas atoms were heavier than the sputtered particles.

Sputtering rates were determined by direct weighing, and although this means long experimental runs for materials of low sputtering rate, particularly at low voltages, it eliminates the dubious interpretation of some other physical quantity in terms of the weight loss.

In the literature several lists of elements are to be found, arranged in their order of sputtering. There is considerable disagreement among the various authors as to these sequences, and it appears impossible to correlate any of them with any of the other physical properties of the elements, or with the periodic table. There appears to be slightly more agreement that the sputtering rate tends to increase with the atomic weight of the bombarding ion. The discrepancies arise owing to the experimental variables discussed in § 3, but the inability to obtain a significant order of sputtering is mainly due to the neglect of secondary electron emission.

§ 2. EXPERIMENTAL TECHNIQUE

2.1. *Discharge Tubes and their Operation*

Figure 1 illustrates a typical tube: side arms support the specimens which constituted the third electrode. A large number of these tubes was constructed.

The cathode was a directly heated oxide coated flat nickel strip supported on an Armco iron base plate and the external connections were made by tungsten leads which passed through steatite insulators in the base plate, and were sealed to the glass foot tube. The various electrodes were shot blasted, degreased and then baked in high vacuum or wet hydrogen where appropriate, before assembly in the tube.

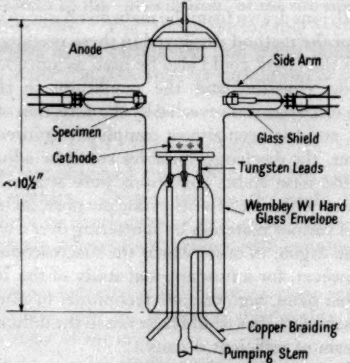


Figure 1. Typical sputtering tube.

Single specimens were mainly used, but occasionally a multiple specimen was made up consisting of six small triangular prisms strapped together to form a hexagonal prism. Before mounting, they were baked under high vacuum where appropriate to drive off water vapour and occluded or adsorbed gases, and were weighed to 10^{-4} g immediately afterwards. The specimen assembly was frequently enveloped in a glass shield as shown in figure 1 : it fitted closely to the contour of the specimen so that, as far as possible, only the face was exposed to bombardment and the supports were shielded. For some tests, specimens without glass shields were placed near the centre of the tube, and the supporting rods were glassed.

After assembly the tube was sealed to a continuous pumping system, and evacuated to a pressure of about 10^{-5} mm Hg. The various electrodes and the glass walls were thoroughly outgassed by heating, the cathode activated and the discharge run at maximum power dissipation until all further outgassing due to ion bombardment had ceased. This was readily detected by the McLeod and Pirani pressure gauges, the colour of the discharge and spectroscopic examination. Thorough outgassing of the specimens was required to ensure stable experimental conditions and the removal of oxide layers and adsorbed gas films, as otherwise sputtering and secondary electron emission representative of the film might have occurred. Outgassing was achieved by raising the specimens to a red heat by positive ion, or electron bombardment, for a short time. Where possible, electron barbardment was preferred, since no corrections for weight loss were then required.

Standard rapid-pumping high vacuum systems, each consisting of one single-stage mercury diffusion pump in series with one three-stage pump and backed by a mechanical pumping system, were employed. For the mercury experiments a small amount of doubly distilled mercury, with less than 0·003% impurities, was placed in each tube before sealing to the pump. For experiments in argon, the gas, with less than 0·01% impurities, was introduced from a glass reservoir through liquid-air traps into the tube. The traps were incorporated between the pumps, the McLeod gauge and the tube to prevent contamination of the argon by mercury vapour and other impurities. During the experiments periodic tests were made spectroscopically of the purity of both the mercury and argon discharges.

For experiments in mercury, the vapour pressure corresponding to the temperature of the coolest part of each tube remained approximately constant and did not exceed about 3×10^{-3} mm Hg. On the other hand, during the argon experiments, the argon 'cleaned up', resulting in a gradual decrease of pressure during bombardment. This was probably due mainly to sputtered atoms trapping argon atoms on the glass walls, but also partly to adsorption of argon on the continuously freshly deposited layers, and to absorption of argon atoms into the glass walls. Similar effects probably occurred with mercury vapour but were unobserved owing to the excess mercury. Thus, for most experiments in argon, fresh argon was admitted at intervals, the pressure limits in each period ranging from 10^{-3} mm to 10^{-2} mm Hg.

The circuit employed is shown in figure 2. The power pack was enclosed in a cage with safety doors and switches, which interlocked with a portable protecting cage surrounding the tube. Voltages and currents were measured by moving-coil instruments to within ± 5%. Bombardment above 8 kv gave rise to x-rays of sufficient intensity for protection with lead and lead glass screens to be necessary. Attention has been drawn to this hazard by the authors (1956).

The current collected by the specimens S was space-charge limited under normal discharge conditions, and could be adjusted independently of the voltage applied, by varying the main discharge current. The limiting resistance R_2 protected the circuit and specimens against instabilities leading to sudden current surges or arcing. Throughout the experiments the specimen current ranged from about 0·1 to 10 mA, and the main discharge current, limited by R_1, from about 0·1 to 15 A. The cathode was normally run at 45 A, 4 V, supplied by the 240 V a.c. main through a step-down transformer. The glass shields round the specimens limited the current drain on the discharge, thus avoiding non-uniformity of the plasma, distortion of the potential distribution and instability. In some of the earlier experiments a rapid comparison of various materials was made with the built up hexagonal multiple specimens. Although a check on the current distribution over the individual prisms showed fairly good uniformity, the later tests were conducted with single specimens.

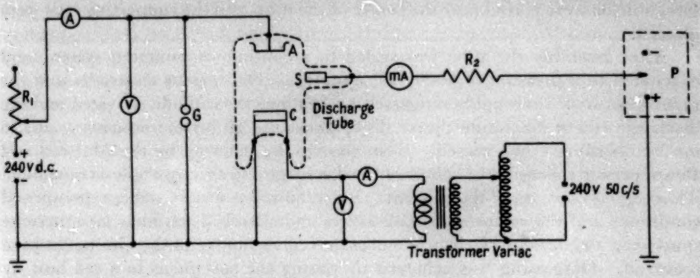


Figure 2. Schematic circuit diagram. R_1, *c.* 20 to 1000 Ω; R_2, *c.* 1 MΩ for work at 500 to 20 000 V, and up to *c.* 1000 Ω for work at under 500 V; P, power pack to deliver smoothed output of up to 20 000 V and 10 mA; G, spark gap to absorb any high voltage flash over from S to A; C, oxide coated cathode taking about 45 A at 4 V; A, anode; S, specimen for ion bombardment.

2.2. *Sputtering Rates and Secondary Electron Emission*

When the experimental runs were completed for each tube, it was sealed off at the constriction while under vacuum. The side arms were cut open and the specimens removed for immediate weighing and examination before contamination by atmospheric conditions could occur. The lowest observable sputtering rates depended on the time allowed for the experiment and the accuracy of the weighing. With carbon this was about 10^{-4} gramme per ampere-hour, although at higher voltages and with other elements this degree of accuracy was unnecessary.

Attempts were made to determine the secondary electron emission from spherical carbon and iron specimens by measuring the diameter of the spherical positive ion sheath round the negatively charged specimen. For these tests each specimen was placed centrally in the plasma and supported by a glassed Nicosel rod which provided the electrical connection. Spherical specimens were chosen to avoid the complications of end effects.

The edge of the sheath was easily sighted against the background of the positive column, especially if it were moved by varying the voltage temporarily. The diameter of the sheath was judged by comparison with the known dimensions of various tube components, the most simple and reliable method tried. Large sheaths were frequently distorted, a condition which could usually be overcome by increasing the discharge current to contract them. From the dimensions, the positive ion current to each specimen was calculated by Langmuir's expressione for space-charge limited currents, and compared with the actual specimen current. This always exceeded the calculated positive ion current except at values below approximately 100 volts, where they agreed. The excess current was assumed to be due to secondary electron emission from the specimen. Owing to distortion of the positive ion sheath at higher voltages, it was not possible to use this method much above 8 kv. The method is discussed in § 4.3.

At 8 kv and above the presence of secondary electrons was proved directly, since x-rays were detected by the fogging, within an hour or so, of dental film placed about six inches from the tube. Their source was determined by pin-hole radiography, and their origin proved to be due to the impact on the walls and other parts of the tube of electrons which arose from the target and were accelerated to high velocities across the positive ion sheath. Occasionally, after continued sputtering and blackening of the tube walls, interesting patterns appeared on the sputtered deposit opposite flat specimens undergoing bombardment. These patterns usually traced the contour of the target and probably arose through local heating of the deposit by electron impact, and its subsequent evaporation. On one occasion, glass microscope cover slips placed opposite a specimen were pitted with small holes, and on another, the tube wall was so heated locally that the glass was softened and 'sucked in'. The x-rays produced were unsuitable for quantitative measurements and subsequent calculation of the secondary electron emission.

A direct method of measuring the secondary electron emission by placing a magnetic field across the face of a flat specimen to return the electrons to the specimen was tried. It failed due to disturbance of the main discharge by the magnetic field. Despite screening, and the application of compensating magnetic fields, the results were unreliable, and the method was abandoned.

§ 3. EXPERIMENTAL RESULTS

The variation of the sputtering rate of polycrystalline carbon, in the form of highly refined spectroscopic graphite, with the energy of the bombarding positive ions of argon and mercury is shown in figure 3. The numbers beside certain experimental points show that the point in question represents the average of that number of experimental determinations. The experimental fluctuations were ± 50% of the mean in some cases, and this method of presentation gives a fair

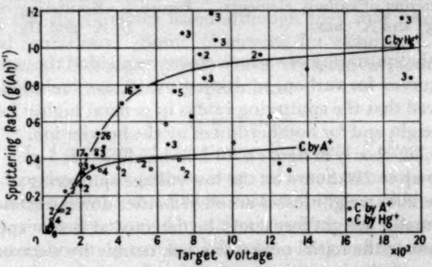


Figure 3. Sputtering of carbon by Hg⁺ and A⁺ ions.

appraisal of the data, as analysis of the variations showed that they fitted a symmetrical distribution curve.

As a result of the relatively intense bombardment of the specimens, the bombarded surfaces, initially flat, were frequently etched to form striking patterns as illustrated in figure 4 (Plate). This surface erosion was probably partly responsible for the experimental variations, as it would inhibit to some extent the escape of sputtered particles. As a variant to the usual type of specimen presenting a plane surface to the bombarding ions, a few experiments were conducted on spherical and cylindrical specimens placed centrally in the discharge. The sputtering rates of these agreed well with the average of the flat specimens, thus indicating that variations in the geometry were relatively unimportant in these experiments. Moreover, the spherical and cylindrical specimens appeared to sputter symmetrically without the formation of patterns, indicating that the most consistent results could probably be obtained from studies on such specimens. In addition, a few tests using tubes of larger dimensions and greater electrode separations than usual supported the conclusion that variations in tube geometry did not influence these results greatly.

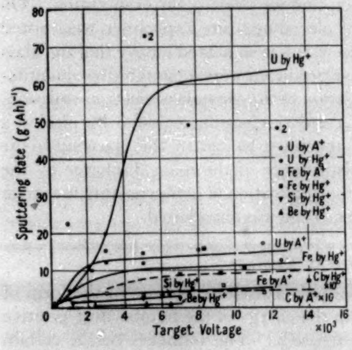


Figure 5. Sputtering of various elements by Hg^+ and A^+ ions.

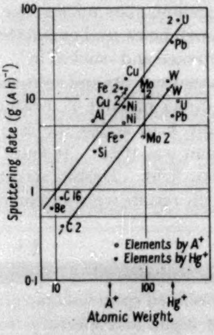


Figure 6. Sputtering of various elements by A^+ and Hg^+ ions at 5 kv.

A survey of the sputtering of various elements yielded the results illustrated in figure 5 : the curves for carbon, multipled by 10, are included for comparison. It will be observed that the sputtering rate is in general higher for the elements of higher atomic weight and for bombardment by the heavier ion. This relationship is brought out more clearly in figure 6 for bombardment at 5 kv.

Tests lasting up to 200 hours on the low voltage sputtering of carbon, iron and uranium gave definite weight losses at ion velocities down to about 30 v, while for carbon no measurable weight loss could be detected at below approximately 15 v. In these experiments the initial outgassing was mainly by electron bombardment. In the few instances where positive ion bombardment was used the weight correction required did not exceed 5% of the measured weight loss. The results are shown in figure 7, where the low ion energies are corrected for the space potential, as determined by probe characteristics. The results for carbon at approximately 10 v refer to specimens held at the floating potential, while that at 15 v refers to an earthed specimen. No attempt was made to take into account either the distribution of ion velocities in the plasma, or the possibility of plasma electron oscillations as indicated by some of the probe characteristics. However the absence of detectable sputtering at 15 v implies that these effects were negligible.

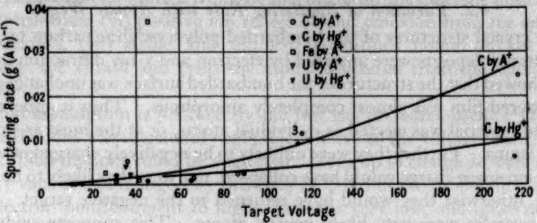


Figure 7. Sputtering of C, Fe and U by low energy Hg^+ and A^+ ions.

The variation of the secondary electron emission from carbon and iron with the energy of the bombarding Hg^+ and A^+ ions is shown in figure 8, where γ = (secondary electron emission)/(positive ion current).

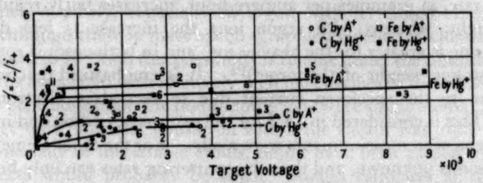


Figure 8. Secondary electron emission from C and Fe by Hg^+ and A^+ ions.

In addition to the variables: 1 ion species and energy, 2 target species, 3 gas pressure, 4 secondary electron emission, 5 surface erosion, 6 tube geometry, experimental studies were carried out on the variation of sputtering with: 7 nominal target current density, 8 target temperature, 9 discharge tube current and 10 duration of experiment. No definite or progressive variations were revealed with the last six variables listed although they may have contributed to the experimental variations. Some observers, for example Campbell (1923), Güntherschulze and Tollmien (1942) have reported evidence of sputtering variations with some of these last six quantities.

§ 4. DISCUSSION

4.1. *Variation of Sputtering Rate with Voltage*

The present experiments on carbon, iron and uranium bombarded by Hg^+ and A^+ and illustrated by figures 3 and 5 show that, with thoroughly outgassed targets whose loss of material is determined by direct weighing, the variation of sputtering rate is approximately linear up to a few kilovolts. Beyond this the rate of increase is gradually reduced, and the relation becomes approximately

$$w = \lambda V^{1/2} \qquad \ldots\ldots(2)$$

where λ is a constant characteristic of the specimen and the bombarding ion. Further, the sputtering rates appear to approach the voltage axis asymptotically indicating that the critical voltages are small. The experiments illustrated by figure 7 show conclusively that sputtering occurs at about 30 v for carbon, iron, and uranium and that the critical voltages for these materials appear to lie between 15 and 30 v.

4.2. *Variation of Sputtering Rate with Atomic Weight*

The crystal structures of the bombarded polycrystalline carbon target and of the sputtered deposits were analysed by electron and x-ray diffraction techniques. These showed that the structure of the bombarded surface was unchanged, whereas the sputtered film was almost completely amorphous. Thus it appears that the sputtered material was ejected as individual atoms, or at the most as small aggregates of atoms. Further they were unlikely to be negatively charged otherwise the positive ion space charge would have collapsed, nor were they likely to be positively charged, otherwise they would have returned to the negative target. Thus the carbon particles were very likely neutral atoms. This supports evidence from magnetic deflection experiments by Baum (1926–27) and from spectroscopic studies by von Hippel (1926) that sputtered particles are neutral atoms.

Hence it seems natural to relate the sputtering rates with the atomic weight as in figure 6. This shows that for bombardment at 5 kv by mercury ions the sputtering rate, in grammes per ampere-hour, increases fairly regularly with the atomic weight, whereas with argon ions the increase is less definite. The sputtering rate is greater for the heavier ion, and in both cases is roughly proportional to (atomic weight of the target)$^{1\cdot 3}$. It is emphasized that in figure 6 the contribution to the specimen current due to secondary electron emission is ignored. This is considered in § 4.5 for the two cases, carbon and iron, for which secondary emission measurements were made. For the other elements the secondary emission is unknown, and thus the sputtering rates can only be expressed in grammes per ampere-hour, instead of the more significant term, atoms per positive ion.

4.3. *Positive Ion Sheaths*

Although the positive ion sheath method of measuring secondary electron emission is indirect, and cannot be regarded as accurate, it does possess the merit of assessment under actual sputtering conditions, without recourse to separate, and perhaps unrepresentative, experimental arrangements.

The space-charge limited positive ion current to the spherical specimens was calculated by the ‘positive ion’ form of the familiar expression due to Langmuir and Compton (1931):

$$i_+ = \tfrac{4}{9}(2e/M)^{1/2}\, V^{3/2}\, \alpha^{-2}\ \text{e.s.u.} \qquad \ldots\ldots(3)$$

where e = charge on ion, M = mass of ion, V = numerical value of negative potential of sphere with respect to the space potential, α^2 is a factor derived mathematically, a function of (radius of sheath)/(radius of sphere).

Values of α^2 are tabulated and presented graphically by Langmuir and Compton (1931), a more convenient alternative graphical representation being given by von Engel and Steenbeck (1934). For Hg^+ and A^+ equation (3) reduced to practical units becomes

$$i_{Hg^+} = 4\cdot 87 \times 10^{-5}\, V^{3/2}\, \alpha^{-2}, \qquad \ldots\ldots(4)$$

and

$$i_{A^+} = 1\cdot 09 \times 10^{-4}\, V^{3/2}\, \alpha^{-2}, \qquad \ldots\ldots(5)$$

where i_+ = positive ion current in mA and V = numerical value of negative voltage across the sheath.

The assumptions implied in the use of Langmuir's equation are that (i) the possibility of collisions between positive ions and atoms in the sheath may be ignored, (ii) the initial velocity of the ions as they enter the sheath may be ignored, (iii) the positive ions are singly charged atoms, (iv) no ionization occurs within the sheath, (v) the electrons emitted from the specimens do not affect the positive space charge distribution, (vi) positive ion reflection and emission from the specimens may be ignored, (vii) stable and uniform conditions obtain in the plasma surrounding the sheath and (viii) the atoms sputtered from the specimen are uncharged.

The first assumption is justified by the low gas pressures used. The initial velocity of the ions was in general negligible relative to the voltages used, and the effects of orbital motions could be neglected. Ions up to charge 5*e* were detected in mercury vapour and in argon by Bleakney (1930), but the probability of ionization by electron bombardment to higher states is very low, and quite negligible for the relatively low energy plasma electrons shown by probe plots to have been present in this work. The secondary electrons were sufficiently energetic to form singly and multiply charged ions, although the probability is so low that, in conjunction with the large electron mean free paths in these experiments, it can be ignored. The formation of negative ions in a discharge through a pure monatomic gas is extremely unlikely, especially in argon with its closed electronic shell structure. Although there is mass spectrographic evidence of cluster formation at pressures down to 10^{-4} mm this possibility is remote under the conditions of large mean free paths associated with relatively high ion energies in these investigations. Thus there is little doubt that the bombarding ions were single atoms carrying one positive charge only, and that ionization within the sheath was unlikely. Some ionization, especially of metastable atoms, might have been caused by radiation, but such effects would probably be small. Rough estimates of the effect of secondary electrons, especially at the surface of the specimen, on the positive ion current, based on Langmuir's (Langmuir 1929, Langmuir and Compton 1931) considerations of double sheaths, indicated that it was negligible. The reflection of positive ions from the negatively charged targets was not likely to be high, as it was very probable that the ions would be neutralized at the surface and, if reflected at all, would be reflected as neutral atoms. The possibility of positive ion emission can be disregarded. During the sputtering runs, the discharge conditions were carefully controlled throughout, so that little trouble was experienced from instability or lack of uniformity of the plasma. That sputtered atoms are unlikely to be charged was discussed in §4·2. Thus it is considered that the theory is valid under our experimental conditions, and that the difference between the calculated positive ion current and the measured specimen current is mainly due to the emission of secondary electrons.

4.4. *Secondary Electron Emission*

Although the possibility of electron emission from sputtered surfaces is realized by some workers it is mainly regarded as insignificant and ignored. However, the present investigations show that secondary electron emission can form the major proportion of the specimen current under certain conditions as illustrated by figure 8.

It is well known that electron emission can arise from an electrode placed in the plasma of a discharge, due to one or more of the following causes: positive ion bombardment, metastable atom bombardment, primary electron bombardment, thermionic emission, photoelectric emission and field emission.

In the present investigations no attempt was made to separate these, but it would appear that the main contributor was the positive ion bombardment. Since the curves of figure 8 tend to pass through the origin, contributions by metastable atoms and photoelectric effects were likely to be small, as, being independent of voltage, they would have resulted in a positive intercept on the $\gamma=i_-/i_+$ axis. In addition, since the secondary electron emission increases only slowly with voltage beyond the knee of the curve; any influences due to the increasing potential gradient and the specimen temperature were slight. Thus neither field nor thermionic emission were likely to be important. The possibility of primary electron bombardment does not arise in these experiments. The emission of negatively charged sputtered atoms was most unlikely, as the blackening of dental film above 8 kv proved conclusively that some at least of the negative particles were electrons, with a strong probability that they were all electrons, as the presence of relatively slow negative atoms near the specimen would have been expected to neutralize the observed positive ion space charge.

The condition of the target surface is known to have a marked effect on the secondary electron emission. These measurements were conducted on thoroughly outgassed, continuously sputtered targets, so that ambiguities due to gas or oxide films were unlikely. Most experiments by other workers on the emission of secondary electrons by positive ion bombardment cover either lower or much higher voltage ranges than that investigated here. The lower voltage experiments generally yield a linear relationship similar to the initial part of the curves of figure 8.

4.5. *Theories*

The two main approaches to the problem of sputtering are the theory of local high temperatures and the theory of direct momentum transfers. The first is based on the idea that the energy of the impinging ion raises the temperature at the point of impact instantaneously to a very high level, so that thermal evaporation occurs. This theory was developed most extensively by von Hippel (1926, 1928), Izmailov (1939), and Townes (1944). The second theory is due mainly to Kingdon and Langmuir (1923) who explain their experiments by postulating that atoms are removed by direct impact. On the other hand experiments by Güntherschulze (1926) led him to propose an electro-chemical mechanism for the sputtering of certain materials, especially in hydrogen discharges. These rival theories are critically examined in the various reviews on sputtering, the most recent of which is by Massey and Burhop (1952).

The main drawback to these theories is that they are all based on experiments where the possibility of secondary electron emission is ignored. Empirical relations such as equations (1) and (2) are useful for practical estimates under similar discharge conditions, but are not so helpful for purposes of theoretical interpretation, as they are merely expressions of the overall macroscopic phenomenon. Since sputtered particles are likely to be neutral atoms a more useful quantity for physical understanding of the individual processes involved is the number of atoms sputtered per positive ion. Let i_+ be the positive ion component of the specimen current and i_- the secondary electron emission, so that the observed current i is i_++i_-. Also let $n=n_++n_-$, the total number of singly charged particles of either sign that strike or are emitted from the specimen per ampere-hour, so that $n = 2{\cdot}25 \times 10^{22}$. Then if N be the number of atoms/gramme of specimen material and s be the sputtering rate in atoms per positive ion striking,

$$s = (N/n_{+})w \quad \text{or} \quad s = (N/n)(1+\gamma)w \qquad \ldots\ldots(6)$$

where $\gamma = i_{-}/i_{+}$ and the ratio N/n is 2·24 and 0·48 for carbon and iron respectively.

The curves of figure 9 for carbon and iron were derived in this way from figures 3, 5 and 8, figure 8 being extrapolated. Figure 9 shows that at first s increases roughly linearly with the ion energy, and later it increases more slowly. The physical interpretation of this effect would seem to be associated with the penetration of the targets by the positive ions: the greater the ion energy, the greater the depth of penetration, the greater the number of collisions, and the more the escape of bombarded atoms is hindered. Thus it would seem that the mathematical problem to be solved is concerned with the manner in which positive ions and bombarded atoms move through solid matter and lose energy, some of the atoms finally escaping.

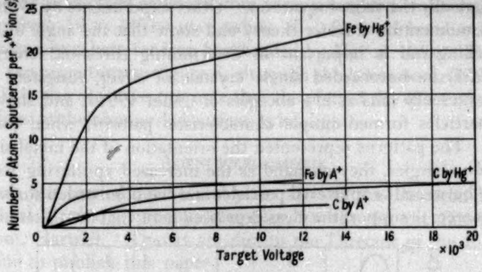


Figure 9. Sputtering of carbon and iron by Hg⁺ and A⁺ ions.

Figure 9 follows, fairly closely, down to a few hundred volts, the empirical relation

$$s = \lambda \ln (V/V_0) \qquad \ldots\ldots(7)$$

where λ = constant characteristic of the gas and the target and V_0 = 'critical' voltage. The values of λ and V_0 are as follows:

Conditions	λ	V_0
Fe by Hg⁺	5·8	270
C by Hg⁺	1·25	430
Fe by A⁺	1·25	150
C by A⁺	0·75	340

V_0 is the intercept at $s = 0$ corresponding to equation (7), and represents approximately the point at which this law breaks down to be replaced by the law represented by figure 7. V_0 may correspond to the point at which the ion has just sufficient energy to 'penetrate' the target. Up to this point 'surface' sputtering occurs. Combining (6) and (7) gives the empirical relation

$$w = (n/N)(1+\gamma)^{-1} \lambda \ln (V/V_0). \qquad \ldots\ldots(8)$$

An interesting suggestion based on neutron diffusion was made by Keywell (1952) which yields a relation of the same form as equation (7). More recently (1955) he extended these ideas and found that sputtering, particularly by light ions, was better described by a radiation damage mechanism. By this the ion can transfer a considerable fraction of its energy to a metallic atom in one collision giving rise to an energetic recoil atom which, in turn, can produce other recoil atoms. These ideas, based on 'penetration' of the target, accord in general with the views put forward in this paper.

Work by Wehner (1954 a) on the threshold energy for sputtering in mercury vapour yields results in disagreement with existing threshold theories (e.g. Kingdon and Langmuir 1923). Wehner related the threshold energy to the bulk sound velocity of the metal. However, his values for carbon, iron and uranium are respectively 500–600, 60–70 and 70–80 volts, which exceed our experimental findings, especially the value for carbon. Other experiments by Wehner (1954 b) support the momentum transfer theory and show that the angle of incidence of the bombarding ion is important in determining threshold energies. More recently (1955) he bombarded single crystals of silver, tungsten, copper and α-iron with mercury ions at low energies of under 150 ev, and showed that the sputtered particles formed simple characteristic patterns when collected on a glass plate. The patterns represented the orientation of the metal single crystal. At higher ion energies, they changed as the increased sputtering led to mutual interference between the sputtered particles and the bombarded surfaces. These experiments seem to support the ideas expressed here, that the material is sputtered as single atoms, that sputtering is essentially a surface phenomenon at low ion energies, and that at higher energies the bombarding ions penetrate below the surface layer of atoms.

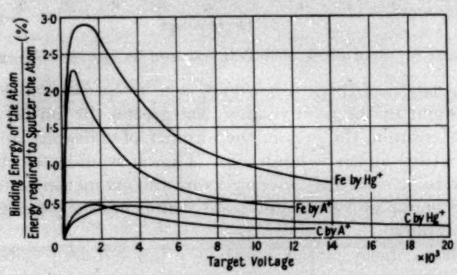


Figure 10. The ratio of the binding energy of the atom to the energy required to sputter it. Carbon and iron bombarded by Hg^+ and A^+ ions.

It is of interest to consider the ratio of the binding energy of the atom to the energy required to sputter it. This ratio is equal to ($s \times$ binding energy)/(target voltage), and the variation with the target voltage is shown in figure 10 for carbon and iron. The values of the binding energies of carbon and iron were calculated from values given by Mott and Jones (1936), and are approximately 6·5 and 4·7 ev per atom respectively at room temperature. The curves illustrate the relatively low values of the ratio, thus demonstrating the small proportion of the available energy that is used in extracting the atom. This again supports the idea of ion penetration and collisions with and between atoms, resulting in the dissipation of most of the ion energy as heat. Further, the existence of a maximum, followed by a rapid decrease in the ratio with energy, can be explained by the increasing depth of ion penetration, remembering however, that part of the original ion energy also imparts velocity to the ejected atoms, and produces secondary electron emission.

These conclusions are reinforced by the general shape of the secondary emission curves of figure 8, which also suggest the penetration of the bombarding positive ions into the surface, as secondaries liberated within the material would be less likely to escape. It is significant in this respect that experiments by Hill *et al.* (1939) ranging from 43 kv to 426 kv for He^+, H_2^+ and H^+ on molybdenum, showed evidence of a maximum followed by a decrease of secondary electron emission with increasing ion energy. This may be compared with the emission of secondaries by penetrating primary electrons. Recent experiments by Bourne *et al.* (1955) on the electron emission from various targets bombarded by positive ions ranging from 10 to 140 kv gave values of γ, for steel bombarded by Hg^+, similar to those given here for iron bombarded by Hg^+ at 10 kv. Further, their curves have the same form as figure 8, although the knees occur at much higher voltages. It is also of interest to note that they found the effect of an extractive voltage gradient at the bombarded surface to be slight.

ACKNOWLEDGMENTS

This work was carried out several years ago at the Research Laboratories of The General Electric Company, Wembley, for the Atomic Energy Research Establishment, Harwell. Thanks are due to the Director of the Establishment for permission to publish this paper.

The authors wish to thank their colleagues of the X-ray and Electron Diffraction Groups at the Research Laboratories for their analytical work. They also wish especially to thank Mr. J. W. Ryde for many stimulating discussions during these investigations.

REFERENCES

BAUM, T., 1926–27, *Z. Phys.*, **40**, 686.
BLEAKNEY, W., 1930, *Phys. Rev.*, **35**, 139; **36**, 1303.
BOURNE, H. C., CLOUD, R. W., and TRUMP, J. G., 1955, *J. Appl. Phys.*, **26**, 596.
CAMPBELL, N. R., *et al.* (Research Staff of the G.E.C.), 1923, *Phil Mag.*, **45**, 98.
COMPTON, K. T., and LANGMUIR, I., 1930, *Rev. Mod. Phys.*, **2**, 123.
VON ENGEL, A., and STEENBECK, M., 1934, *Elektrische Gasentladungen*, Vol. 2 (Berlin: Springer).
GÜNTHERSCHULZE, A., 1926, *Z. Phys.*, **36**, 563; 1955, *Ibid.*, **141**, 346.
GÜNTHERSCHULZE, A., and TOLLMIEN, W., 1942, *Z. Phys.*, **119**, 685.
HILL, A. G., BUECHNER, W. W., CLARK, J. S., and FISK, J. B., 1939, *Phys. Rev.*, **55**, 463.
VON HIPPEL, A., 1926, *Ann. Phys., Lpz.*, **80**, 672; 1928, *Ibid.*, **81**, 1043., **86**, 1006.
IZMAILOV, S. V., 1939, *Zh. Exp. Teor. Fiz.*, **9**, 1469.
KEYWELL, F., 1952, *Phys. Rev.*, **87**, 160; 1955, *Ibid.*, **97**, 1611.
KINGDON, K. H., and LANGMUIR, I., 1923, *Phys. Rev.*, **22**, 148.
LANGMUIR, I., 1929, *Phys. Rev.*, **33**, 954.
LANGMUIR, I., and COMPTON, K. T., 1931, *Rev. Mod. Phys.*, **3**, 191.
MASSEY, H. S. W., and BURHOP, E. H. S., 1952, *Electronic and Ionic Impact Phenomena* (Oxford: University Press).
MEYER, K., and GÜNTHERSCHULZE, A., 1931, *Z. Phys.*, **71**, 279.
MOTT, N. F., and JONES, H., 1936, *The Theory of the Properties of Metals and Alloys* (Oxford: University Press).
OLIPHANT, M. L., 1928, *Proc. Camb. Phil. Soc.*, **24**, 451.
RESEARCH STAFF OF THE G.E.C., LONDON, 1923, *Phil. Mag.*, **45**, 98.
STRACHAN, J. F., and HARRIS, N. L., 1956, *Nature Lond.*, **178**, 588.
TOWNES, C. H., 1944, *Phys. Rev.*, **65**, 319.
WEHNER, G. K., 1954 a, *Phys. Rev.*, **93**, 633; 1954 b, *J. Appl. Phys.*, **25**, 270; 1955, *Ibid.*, **26**, 1056.

RESEARCH NOTES

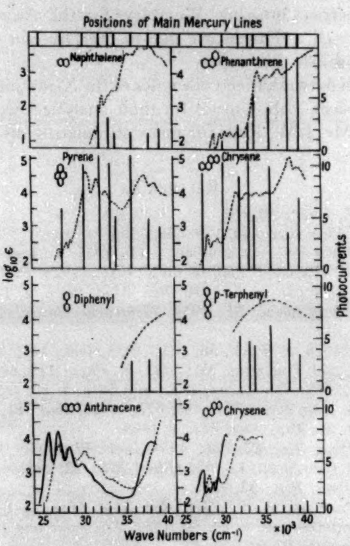
Photo- and Semi-Conduction of Aromatic Hydrocarbon Crystals

By L. E. LYONS and G. C. MORRIS

Department of Physical Chemistry, University of Sydney, Australia

MS. received 15*th May* 1956

Recent work (e.g. Putseiko 1949, Akamatu and Inokuchi 1950, Vartanyan 1950, Inokuchi 1951, Eley *et al.* 1953, Mette and Pick 1953, Carswell 1953, Bree, Carswell and Lyons 1955, Northrop and Simpson 1954, Chynoweth and Schneider 1954, Lyons 1955) has confirmed that both photo- and semi-conduction take place in some aromatic hydrocarbon crystals. We now report the spectral dependence of the photocurrent in a series of compounds (see figure) including naphthalene, phenanthrene, diphenyl and p-terphenyl, all of which have not previously been shown to be photoconductors. With the



Crystal photocurrents and solution absorption spectra. (The photocurrent scale on the extreme right applies to all graphs, but the units differ from one graph to another. ϵ, the molecular extinction coefficient.)

light intensity used we could not detect a photocurrent in hexamethyl benzene nor in 1, 2, 3, 4, 5, 6, 7, 12, 13, 16-decahydropyrene. The results are shown for the various lines obtained from a mercury lamp corrected for intensity variation. They refer to single crystals in dry air. Also shown in the figure are the solution spectra of the compounds (dotted lines) taken from Friedel and Orchin (1951).

The absorption by the crystals is expected in very nearly the same wavelength region as for solutions. It is seen that photocurrents are obtained over the absorption region but not outside it. Only in the cases of chrysene and anthracene was excitation with a continuum sufficiently intense to give a detectable photocurrent, and in these cases the resemblance between the photocurve and the absorption spectrum is most marked, as with tetracene (Bree and Lyons 1954). Anthracene has been discussed fully (Carswell and Lyons 1955) and chrysene has also been reported to give a photocurrent between 3800Å and 5400Å (Northrop and Simpson 1955). A current at wavelengths above 4000Å was obtained by us but was traced to the effect of stray light in the monochromator.

The correlation (Carswell, Ferguson and Lyons 1954) between electronic transitions and photocurrents is confirmed. It is possible that in photoconduction the excited singlet level may be an intermediate in the formation of a triplet level, the importance of which in semiconduction was suggested by the same authors. Northrop and Simpson (1956) have since found evidence in mixed crystals which confirms the part played by the triplet level. If the triplet level is also important in photoconduction the light and heat effects are closely related.

The order of magnitude of the photocurrent I is of interest. When the experiments were carried out under comparable conditions the following values for $-\log I$ were obtained as an average of the observed currents at the various wavelengths used within the absorption region: Diphenyl, 16; *p*-terphenyl, 15·3; naphthalene 13·5; phenanthrene, 13·0; pyrene, 12·3; chrysene, 11·5; anthracene 11·3; tetracene, 9·3. The arrangement in order of the substances is similar irrespective of which of the following properties is considered: the magnitude of the photocurrent; the energy of an electronic transition from the molecular ground state to the lowest excited singlet (or triplet) level; the ionization potential of the molecule; the electron affinity of the molecule; the maximum free valence in the molecule; the minimum 'atom localization' energy; the radical (methyl or trichloromethyl) affinity. A full discussion of these correlations will be presented elsewhere. It is obvious that the current depends upon properties of the molecule. The rule (Bree, Carswell and Lyons 1955) that the current is greater the more easily the compound is oxidized, is consistent with the present results.

References

Akamatu, H., and Inokuchi, H., 1950, *J. Chem. Phys.*, **18,** 810.

Bree, A., Carswell, D. J., and Lyons, L. E., 1955, *J. Chem. Soc.*, 1728.

Bree, A., and Lyons, L. E., 1954, *J. Chem. Phys.*, **22,** 1630.

Carswell, D. J., 1953, *J. Chem. Phys.*, **21,** 1890.

Carswell, D. J., Ferguson, J., and Lyons, L. E., 1954, *Nature, Lond.*, **173,** 736.

Carswell, D. J., and Lyons, L. E., 1955, *J. Chem. Soc.*, 1734.

Chynoweth, A. G., and Schneider, W., 1954, *J. Chem. Phys.*, **22,** 1021.

Eley, D. D., Parfitt, G. D., Perry, M. J., and Taysum, D. H., 1953, *Trans. Faraday Soc.*, **49,** 79.

FRIEDEL, R. A., and ORCHIN, M., 1951, *Ultraviolet Spectra of Aromatic Compounds* (New York: Wiley).
INOKUCHI, H., 1951, *Bull. Chem. Soc., Japan*, **24**, 222.
LYONS, L. E., 1955, *J. Chem. Phys.*, **23**, 220.
METTE, H., and PICK, H., 1953, *Z. Phys.*, **134**, 566.
NORTHROP, D. C., and SIMPSON, O., 1954, *Proc. Phys. Soc.* B, **67**, 892; 1955, *Ibid.*, **68**, 974; 1956, *Proc. Roy. Soc.* A, **234**, 124.
PUTSEIKO, E. K., 1949, *Doklady Akad. Nauk., S.S.S.R.*, **67**, 1009.
VARTANYAN, A., 1950, *Doklady Akad. Nauk, S.S.S.R.*, **71**, 641.

LETTERS TO THE EDITOR

A Stabilized Retinal Image of the Iris

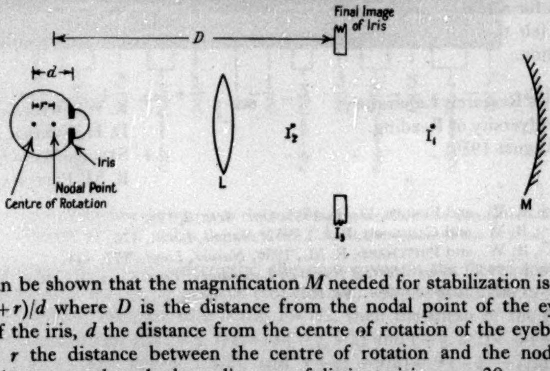
It is now well established that small involuntary eye movements take place when a subject fixates as carefully as possible on a stationary point target. These movements result in a scanning motion of the image across the retinal receptors. In an endeavour to assess the possible functional significance of these movements various optical systems have been devised (Ditchburn and Ginsborg 1952, Ratliff 1952, Ditchburn and Fender 1955, Ditchburn and Pritchard 1956, Krauskopf 1956) to keep the image stationary on the retina despite eye movements.

All of these systems for producing a stabilized retinal image require the subject to wear a close fitting contact lens. This severely restricts the number of subjects available and has resulted in the work being carried on by three small research groups in each of which the role of subject and experimenter is interchangeable.

Most of the measurements made in this work depend on subjective judgments and it is, therefore, desirable that some, at least, of the results should be checked by experiments with a large number of subjects who are not members of one of the research teams.

The cost of providing a large number of subjects with contact lenses is prohibitive. The method described below does not require the subject to wear a contact lens and does not demand any preliminary training of the subject.

The subject, using monocular vision, looks into an optical system as shown in the diagram and sees therein an image of his own iris; this iris, of course, moves with the eyeball. The magnification of the optical system is such that the image of the iris makes the same angular excursion as the eye itself and consequently the image of the iris on the retina is stabilized.



It can be shown that the magnification M needed for stabilization is given by $M=(D+r)/d$ where D is the distance from the nodal point of the eye to the image of the iris, d the distance from the centre of rotation of the eyeball to the iris and r the distance between the centre of rotation and the nodal point. D must be greater than the least distance of distinct vision, say 30 cm, r is about 8 mm for most eyes and thus is small compared with D, but d varies from 1·1 to 1·3 cm for different subjects; thus the magnification must be about ×25 and adjustable to suit the subject. The lens L forms a real magnified image of the iris at I_1, the concave mirror M forms a further magnified real image of I_1 at I_2. The subject sees a virtual image of I_2 formed at I_3 by the light on its second transit through the lens. The magnification at each stage is arranged to be approximately the same; the focal lengths and positions of the components are chosen so that the overall magnification is the value required for stabilization. The variation in magnification to accommodate different subjects can be achieved by altering the position of the components slightly. This is done by moving them along an optical bench by means of lead screws of suitable pitch.

The apparatus does not stabilize against head movements. It is, therefore, essential that the subject's head should be firmly fixed. It must also be so placed that the iris lies accurately on the optic axis of the system, for a misalignment of only 1 mm will move the image about 25 mm off axis.

The opportunity of demonstrating this method of stabilizing a retinal image arose at the Physical Society Exhibition in May 1956. It was obviously not possible to provide a rigid head support to fit every visitor to the stand; however, a moulded headband capable of horizontal adjustment and a chin rest for adjustment in the vertical direction afforded a certain amount of assistance in this respect.

The subject's head was brought to the correct position by moving the head and chin rests until the subject looked through a 'peephole' on the optic axis of the system; the peephole was then removed. A remaining error equal to the radius of the pupil was still possible, however, and precise adjustments required the co-operation of the subject. Finally, the head rest was moved along the optical bench until the iris was at the correct distance from the lens.

It is estimated that about 300 persons attempted to view the stabilized image of their own iris—about half of these were able to 'find the image'. Over 95% of those who located the image reported some or all of the effects which have been previously observed with a stabilized image, viz. loss of contrast discrimination, fading of the image to a uniform grey field and, in some cases, a total loss of vision for a few seconds.

It is felt that these results supplement those obtained by other methods of stabilization.

Physics Research Laboratory,
University of Reading.
3rd August 1956.

R. W. Ditchburn.
D. H. Fender.
Stella Mayne.
R. M. Pritchard.

Ditchburn, R. W., and Fender, D. H., 1955, *Opt. Acta*, **2**, 128.
Ditchburn, R. W., and Ginsborg, B. L., 1952, *Nature, Lond.*, **170**, 36.
Ditchburn, R. W., and Pritchard, R. M., 1956, *Nature, Lond.*, **177**, 434.
Krauskopf, J., 1956, *Army Medical Research Laboratory*, Fort Knox, Ky, Rep. No. 221.
Ratliff, F., 1952, *J. Exp. Psychol.*, **43**, 163.

Paramagnetic Resonance Absorption in ^{235}UCl$_3$, and the Nuclear Electric Quadrupole Moment of ^{235}U

Paramagnetic resonance absorption in single crystals of (La, U)Cl$_3$ containing ^{235}U in its normal abundance has been examined recently by Hutchison, Llewellyn, Wong and Dorain (1956). Crystals of LaCl$_3$ containing about 0·1 mole per cent of UCl$_3$ enriched in ^{235}U have now been prepared, which give a greatly improved signal/noise ratio on the hyperfine lines. This has made it possible to investigate electronic transitions in which the nuclear magnetic quantum number *m* changes by ±1 or ±2. These lines have been measured at 20°K using wavelengths of 3·2 and 1·3 cm, and the magnitude of the nuclear electric quadrupole interaction can be deduced from them. The spectrum has been analysed using the spin Hamiltonian ($S=1/2$, $I=7/2$)

$$\mathscr{H} = \beta\{g_\parallel H_z S_z + g_\perp(H_x S_x + H_y S_y)\} + A S_z I_z + B(S_x I_x + S_y I_y) \\ + P\{I_z^2 - \tfrac{1}{3}I(I+1)\} - g_N \beta_N H I \qquad \ldots\ldots(1)$$

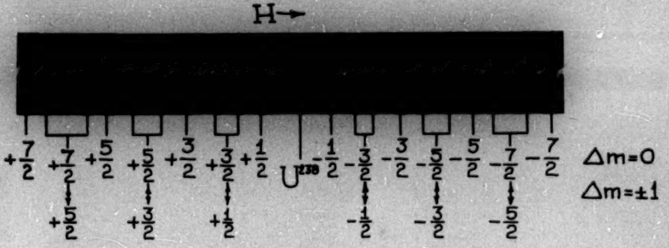
where the parameters

$$g_\parallel = 4{\cdot}153 \pm 0{\cdot}005, \qquad |A| = 0{\cdot}0176 \pm 0{\cdot}001\ \mathrm{cm}^{-1}$$
$$g_\perp = 1{\cdot}520 \pm 0{\cdot}002,$$

are as measured previously by Hutchison *et al.* (1956). We now obtain $|B| = (58{\cdot}5 \pm 0{\cdot}5) \times 10^{-4}\ \mathrm{cm}^{-1}$, $|P| = (5{\cdot}5 \pm 0{\cdot}5) \times 10^{-4}\ \mathrm{cm}^{-1}$.

The figure shows the spectrum observed at a wavelength of 3·24 cm when the applied field *H* makes an angle of 70·2° with the *c* axis of the crystal. The stronger lines correspond to transitions in which $\Delta m = 0$, and the weaker lines to transitions in which $\Delta m = \pm 1$. The oscilloscope trace represents the derivative of the absorption curve versus field strength; the signs given in the figure correspond to the assumption that g_N is positive (see below). If the formulae of Bleaney (1951) are used for the case of $S=1/2$, $I=7/2$, then the mean of the intervals between the transitions, (*a*) $(+\frac{1}{2}, 7/2) \longleftrightarrow (-\frac{1}{2}, 5/2)$ and $(+\frac{1}{2}, 5/2) \longleftrightarrow (-\frac{1}{2}, 7/2)$, and (*b*) $(+\frac{1}{2}, -7/2) \longleftrightarrow (-\frac{1}{2}, -5/2)$ and $(+\frac{1}{2}, -5/2) \longleftrightarrow (-\frac{1}{2}, -7/2)$, which correspond to the two weaker lines at each end of the spectrum, is given by

$$6|P|\left\{\frac{3A^2 g_\parallel^2}{K^2 g^2}\cos^2\theta - 1\right\}$$



Oscilloscope trace of the first derivative of the paramagnetic resonance absorption at 3·42 cm wavelength when *H* makes the angle 70·2° with the *c* axis of the crystal.

where $K^2g^2 = A^2g_{\parallel}^2\cos^2\theta + B^2g_{\perp}^2\sin^2\theta$, and terms of the order P^3/B^2 are omitted. Measurements at the angle and wavelength corresponding to the figure give $|P| = 5{\cdot}8\times10^{-4}\,\mathrm{cm}^{-1}$ while similar measurements at $\theta = 75{\cdot}1°$ and 1·3 cm wavelength give $|P| = 5{\cdot}5\times10^{-4}\,\mathrm{cm}^{-1}$.

When the external magnetic field H is normal to the c axis of the crystal, lines corresponding to transitions in which $\Delta m = \pm 2$ are observed together with lines for which $\Delta m = 0$. Using the formulae of Bleaney (1951) extended to include terms of order P^3/B^2 we find that in this orientation the mean of the two intervals between the transitions (*a*) $(+\tfrac{1}{2}, 7/2)\longleftrightarrow(-\tfrac{1}{2}, 3/2)$ and $(+\tfrac{1}{2}, 3/2)\longleftrightarrow(-\tfrac{1}{2}, 7/2)$, and (*b*) $(+\tfrac{1}{2}, -7/2)\longleftrightarrow(-\tfrac{1}{2}, -3/2)$ and $(+\tfrac{1}{2}, -3/2)\longleftrightarrow(-\tfrac{1}{2}, -7/2)$ is given by the quantity

$$\left|10P - \frac{105P^3}{B^2-25P^2} + \frac{30P^3}{B^2-P^2}\right|.$$

For the similar mean involving the $m = \pm 5/2$ and $m = \pm\tfrac{1}{2}$ levels we have

$$\left|6P - \frac{135P^3}{B^2-9P^2} - \frac{30P^3}{B^2-P^2}\right|.$$

The value assumed for B in the calculation of P from these expressions is not critical, and they lead to the values $5{\cdot}6\times10^{-4}$ and $5{\cdot}8\times10^{-4}\,\mathrm{cm}^{-1}$ respectively for $|P|$, from measurements at 1·3 cm wavelength. From similar measurements at 3·2 cm, which are subject to somewhat larger uncertainties, we find $|P| = 4{\cdot}9\times10^{-4}\,\mathrm{cm}^{-1}$.

In the same orientation with H normal to the c axis, the separation of the two extreme stronger lines arising from $m = \pm 7/2$ transitions is given by

$$\left|7B + \frac{21P^2B}{B^2-25P^2}\right|.$$

The separation of the corresponding pair of lines arising from the $m = \pm 5/2$ transitions is given by

$$\left|5B + \frac{45P^2B}{B^2-9P^2}\right|.$$

Assuming $|P| = 5{\cdot}5\times10^{-4}\,\mathrm{cm}^{-1}$ we find for $|B|$ the values $58{\cdot}2\times10^{-4}$ and $58{\cdot}5\times10^{-4}\,\mathrm{cm}^{-1}$ respectively using measurements at 1·3 cm wavelength, and $59{\cdot}3\times10^{-4}\,\mathrm{cm}^{-1}$ from a measurement at 3 cm.

The value of the nuclear electric quadrupole moment may be deduced from $|P|$ using the formulae of Elliott and Stevens (1953). We shall estimate values of $\langle r^{-3}\rangle$ from the spin–orbit coupling energy ζ in the same way as Bleaney (1955), though this probably leads to too large a result for 5f electrons. The values of $1300\,\mathrm{cm}^{-1}$ for ζ given by Eisenstein and Pryce (1955) for $\mathrm{NpO_2}^{2+}$, and $1700\,\mathrm{cm}^{-1}$ given by Jörgensen (1955) for U^{3+}, thus lead to values of $46A^{-3}$ and $56A^{-3}$ respectively. Taking a mean value of $50A^{-3}$, which corresponds to a nuclear magnetic moment of 0·35 n.m. from the data given above for A, B, we obtain $|Q| = 3{\cdot}8\times10^{-24}\,\mathrm{cm}^2$, with no correction for polarization of the electron core. Apart from the considerable uncertainty in $\langle r^{-3}\rangle$, there is an error of perhaps 10% in our application of the formulae of Elliott and Stevens, which arises through admixture of higher multiplet levels by the crystal field.

From some small asymmetries in the observed spectrum it appears probable that g_N (and hence also B) is positive, and that P is negative; Q is then also negative.

We wish to thank the Atomic Energy Research Establishment, Harwell, for the loan of enriched ^{235}U, and Drs. P. Dorain, E. Wong and G. Candela of the University of Chicago for preparation of the $LaCl_3$ powder from which our crystals were grown. One of us (C. A. H.) is indebted to the Clarendon Laboratory for the privilege of visiting and working there.

Clarendon Laboratory,
Oxford.
27th July, 1956.

B. Bleaney.
C. A. Hutchison, Jr.†
P. M. Llewellyn.
D. F. D. Pope.

† Fellow of the John Simon Guggenheim Foundation.

References

Bleaney, B., 1951, *Phil. Mag.*, **42,** 441; 1955, *Proc. Phys. Soc.* A, **68,** 937.
Eisenstein, J. C., and Pryce, M. H. L., 1955, *Proc. Roy. Soc.* A, **229,** 20.
Elliott, R. J., and Stevens, K. W. H., 1953, *Proc. Roy. Soc.* A, **218,** 553.
Hutchison, C. A., Jr, Llewellyn, P. M., Wong, E., and Dorain, P., 1956, *Phys. Rev.*, **102,** 292.
Jörgensen, C. K., 1955, *K. Danske Videnskab. Selskab, Mat.-fys. Medd.*, **29,** No. 11, 1.

The Paramagnetic Resonance Spectrum of Fe^{3+} in the Cubic Field of MgO †

The ferric ion Fe^{3+} has $3d^5$ electrons, a half-filled shell, and therefore a ground state of $^6S_{5/2}$. In a cubic crystalline electric field the six-fold degeneracy is lifted and is split into a two-fold and four-fold degenerate level (Bethe 1929). Van Vleck and Penney (1934) have discussed the mechanism by which such a splitting can be effected. The splitting is presumably caused by higher order interaction involving the cubic field and spin–orbit coupling. The Hamiltonian can be written as:

$$\mathscr{H} = g\beta \mathbf{H} \,.\, \mathbf{S} + \tfrac{1}{6}a\{S_x{}^4 + S_y{}^4 + S_z{}^4 - \tfrac{1}{5}[S(S+1)(3S^2+3S-1)]\} \ldots\ldots(1)$$

where $3a$ is the separation between the two-fold and four-fold degenerate levels. The transitions are given by

$$\left.\begin{array}{ll} M = \tfrac{1}{2} \rightarrow -\tfrac{1}{2} & H = H_0 + \epsilon_3 \\ M = \pm \quad \rightarrow \pm\tfrac{1}{2} & H = H_0 \pm \tfrac{5}{2}pa + \epsilon_2 \\ M = \pm\tfrac{5}{2} \rightarrow \pm\tfrac{3}{2} & H = H_0 \pm 2pa + \epsilon_1 \end{array}\right\} \ldots\ldots(2)$$

$p = (1-5\phi)$, $\phi = 1^2m^2 + 1^2m^2 + m^2n^2$; (l, m, n) being the direction cosines of the magnetic field with respect to the cubic axes; ϵ_1, ϵ_2, ϵ_3 are higher order terms involving $a^2/g\beta H$ and are given explicitly by Bleaney and Trenam (1954). Neglecting the higher order terms the spectrum is characterized by a line at $g = 2$ (approximately) flanked by two satellites at each side with separations in the ratio of 4 : 5. The relative intensities of this quintuplet are 8 : 5 : 9 : 5 : 8.

We have observed the paramagnetic resonance spectrum of Fe^{3+} in cubic Mg0 at 1·2 cm wavelength at room and liquid nitrogen temperatures. The total amount of iron in the sample was about 0·1% by weight, a large part of which was in the ferrous state. The spectrum can be described by the above Hamiltonian

---

† This work was supported by the U.S. Atomic Energy Commission.

(equation (1)) with the parameters $g = 2{\cdot}0037 \pm 0{\cdot}0007$ and $3a = +615 \times 10^{-4}\,\mathrm{cm}^{-1}$. The positive sign of *a* was determined from the relative intensities of the fine structure lines at liquid helium temperature.

These results are significant in two respects:

(1) The cubic splitting is considerably larger than that found in various alums investigated by Bleaney and Trenam (1954). This points to a much stronger cubic field. A similar increase in the cubic field splitting has been noticed in Mg0 : Mn for which *a* is also about 2–2·5 times larger than in hydrated Tutton salts (Low 1956). In part, this increase in the cubic field may be caused by the smaller interatomic distance in Mg–0 compared with the alums. The distance may be even smaller in the case of the ferric ion because of the electrostatic attraction caused by the excess charge on the ferric ion. This will result in considerable π bonding. Evidence has been accumulated that π bonding is a very important factor in the interpretation of various 3d ions in Mg0 (Low, to be published).

It is significant that the sign of *a* is positive, whereas in the various alums a negative sign is found. A similar positive *a* is found for Mn^{2+} in the same crystal. It is also confirmed now that the *g* factor for Fe^{3+} is larger than that of Mn^{2+} and larger than that of the free electron. It is probable th[illegible] explanation of these differences will depend on an accurate calculation o[illegible] excited states (^4P, ^4D, ^4G, etc.).

The ratio of the cubic splittings $a_{Fe^{3+}}/a_{M}$[illegible] approximately 11 : 1 in MgO. Similar ratios were found for various hydrated complexes. This indicates that this ratio [illegible] only slightly dependent on the crystalline environment. It is highly suggestive that the much larger value of the cubic splitting in Fe^{3+} is caused by the larger spin–orbit coupling.

(2) The spectrum is cubic despite the fact that the paramagnetic trivalent ion [illegible] a divalent lattice. If there were a[illegible]ciated with vacancies the cubic symmetry would be destroyed. The meas[illegible]ensity ratio indicates some disorder in the crystal, possibly because of randomly distributed vacancies or excess oxygen ions. The measured intensities are 3·8 : 1·4 : 9 : 1·4 : 3·8 and the ratio of the line widths 1·5 : 2·1 : 1 : 2·1 : 1·5. The ratio of the integrated intensity of the two satellite lines is about 2·7 : 1. Similar anomalies were found for Mn^{2+} in the very same crystal. Randomly distributed defects would effect the $^{2} \to \frac{3}{2}$ transition more than the other transitions and reduce the intensities and increase the line width. Another possibility is that covalent bonding might change the theoretical intensity ratio. Full details will be published at a later date.

The author gratefully acknowledges the friendly encouragement of Professor C. A. Hutchison, Jr., and the instructive correspondence of Dr. B. Bleaney.

Enrico Fermi Institute for Nuclear Studies,
University of Chicago, Chicago, Illinois. W. Low.

30th July 1956.

Bethe, H., 1929, *Ann. Phys., Lpz.*, **3**, 133.
Bleaney, B., and Trenam, S., 1954, *Proc. Roy. Soc.* A, **223**, 1.
Low, W., 1956, *Phys. Rev.*, **101**, 1827.
van Vleck, J. H., and Penney, W. G., 1934, *Phil. Mag.*, **17**, 961.

REVIEWS OF BOOKS

Instruments for Measurement and Control, by W. G. Holzbock. Pp. vi + 371. (New York: Reinhold Publishing Corporation; London: Chapman and Hall, 1955.) 80*s*.

The author of this book is a research engineer with the Askania Regulator Company of Chicago. The book has therefore an ' engineering ' approach to the field of instrumentation in industry, and will be of direct interest to the industrial rather than to the academic physicist. There is hardly a formula in the book and almost no mathematics, even in the section dealing with control. Instead the author has concentrated upon the description of a wide range of existing instruments used for industrial measurements, and the elements such as pneumatic valves by which control of an industrial process is exercised. There are separate chapters dealing with measurements of temperature, humidity, pressure, flow, liquid level, density, viscosity, speed, and of the physical properties such as absorption spectra by which a chemical process can be followed.

It will fascinate the physicist to see how many physical effects have been used in industrial measurements. Liquid level has been observed in opaque tanks, for example, by gamma rays, and liquid flow has been measured industrially by allowing the liquid to flow through a flexibly mounted loop, and measuring the gyroscopic effect on the loop. The author goes astray in his efforts to explain the gyroscopic effect in simple terms, but the merit of the book is that it provides what appears to be a comprehensive survey of the many physical effects that have been successfully used.

Although this is not the only book on the subject, it will also serve to bring to the notice of the physicist the precision which has been achieved in pneumatic controllers, based on the gauging of the distance of a jet nozzle from a plane surface by the back pressure in the nozzle. One micron displacement is easily detected. Incidentally, the author thinks that pneumatic control systems succeeded in America partly because of their intrinsic safety in the explosive atmosphere of some chemical plants. He thinks that electric controllers might have fared better had the Americans adopted the standards of intrinsic safety in electrical equipment that exist in Great Britain, due to the painstaking work of the Safety in Mines organization, and he considers that the future may still lie with the electric controller.

The book is outstanding for its diagrams; they are well drawn and well reproduced. Moreover they are extremely generously provided—there are less than twenty places at which the book can be opened without at least one diagram being visible. This is a stimulating book which can be read with interest by any physicist who wishes to see something of the ingenuity of the control engineer in exploiting physical effects, and its pages may well suggest an answer for the reader who has a problem in industrial measurement.

R. V. JONES.

CORRIGENDUM

The Absorption of Ultrasonic Waves in a Number of Pure Liquids over the Frequency Range 100 *to* 200 *Mc/s*, by E. L. Heasell and J. Lamb (*Proc. Phys. Soc.* B, 1956, **69**, 869).

The entry in the table for water (no. 92) is wrong; the correct values are:

	Liquid	T_3	f_3	α_3/f_3^2
92.	Water (distilled)	24.7	192.8	21.4

CONTENTS OF SECTION A

PAGE

Mr. M. J. **Laubitz**. Pseudoscalar Interaction in the Decay of ^{144}Pr 789

Dr. M. H. **Alston**, Dr. W. H. **Evans**, Dr. M. **Fidecaro**, Mr. R. W. **Newport**, Dr. G. **von Gierke** and Mr. P. R. **Williams**. π^+-p Scattering at 10–35 mev in a Diffusion Cloud Chamber 798

Mr. R. J. **Norman**. The Energy Spectrum of Extensive Air Showers . . 804

Dr. A. **Dalgarno** and Mr. N. **Lynn**. Resonance Forces at Large Separations . 821

Mr. M. G. **Rusbridge**. On the Reaction ^{13}C(α, n)^{16}O 830

Mr. W. **Stanners** and Dr. M. A. S. **Ross**. The β-disintegration of RaD . . 836

Dr. P. D. **McCormack**. A Liquid Scintillator Investigation of the Second Maximum in the Cosmic Ray Transition Curve 845

Mr. G. N. **Harding** and Dr. F. J. M. **Farley**. The Angular Distribution of Neutrons from High Energy Fission. 853

Research Notes :

Mr. R. **Berenbaum**, Mr. J. H. **Towle** and Mr. J. H. **Matthews**. The 440 kev Transition in ^{23}Na 858

Prof. W. E. **Burcham** and Mr. B. C. **Haywood**. The Alpha Activity Induced in Platinum by Bombardment with Nitrogen Ions 862

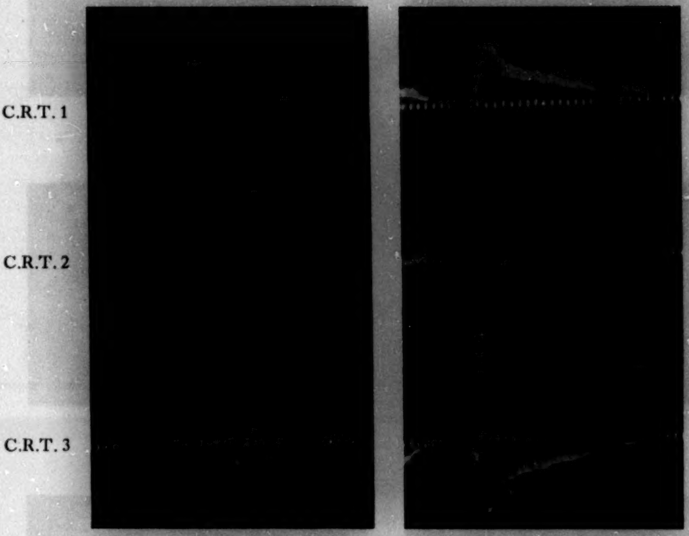
Mr. J. **Steyn**. Absolute Standardization of Beta-emitting Isotopes with a Liquid Scintillation Counter 865

Dr. A. C. **Hurley**. On the Binding Energy of the Helium Hydride Ion. . 868

Dr. J. R. **Prescott**. The Density Distribution of Extensive Showers . . 870

Reviews of Books 875

Contents of Section B 876



(*a*) (*b*)

Figure 1. Film record. (*a*) Echo showing high wind velocity. (*b*) Echo showing low wind velocity with phase change produced by resonance (X to Y).

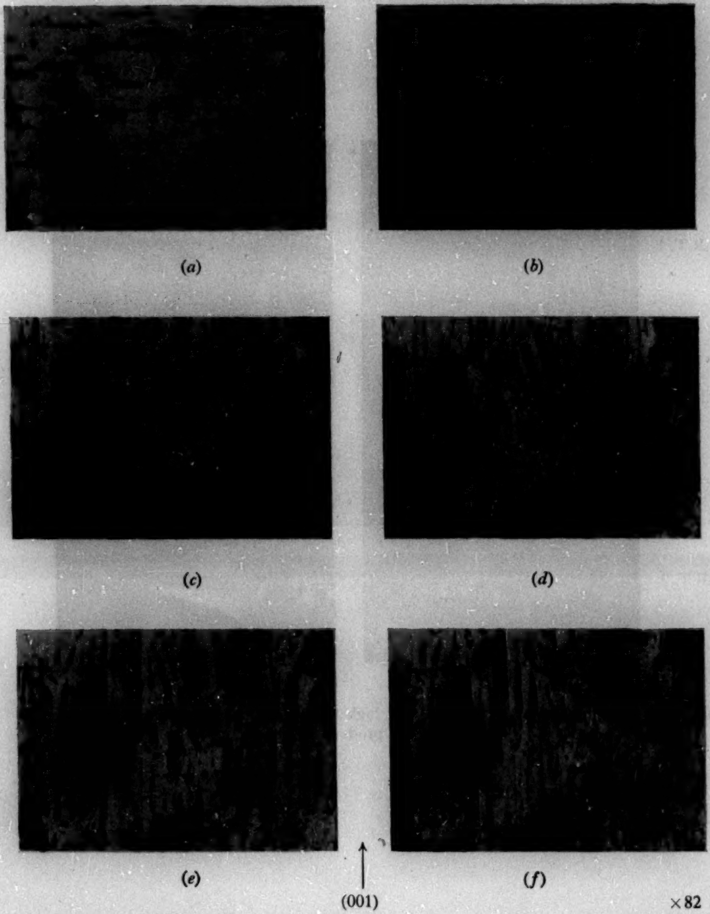


Figure 2. Appearance of domain patterns on (110) surface of silicon–iron; field decreasing from saturation value (*a*) to zero value at (*d*); field increasing in reverse direction in (*e*) and (*f*).

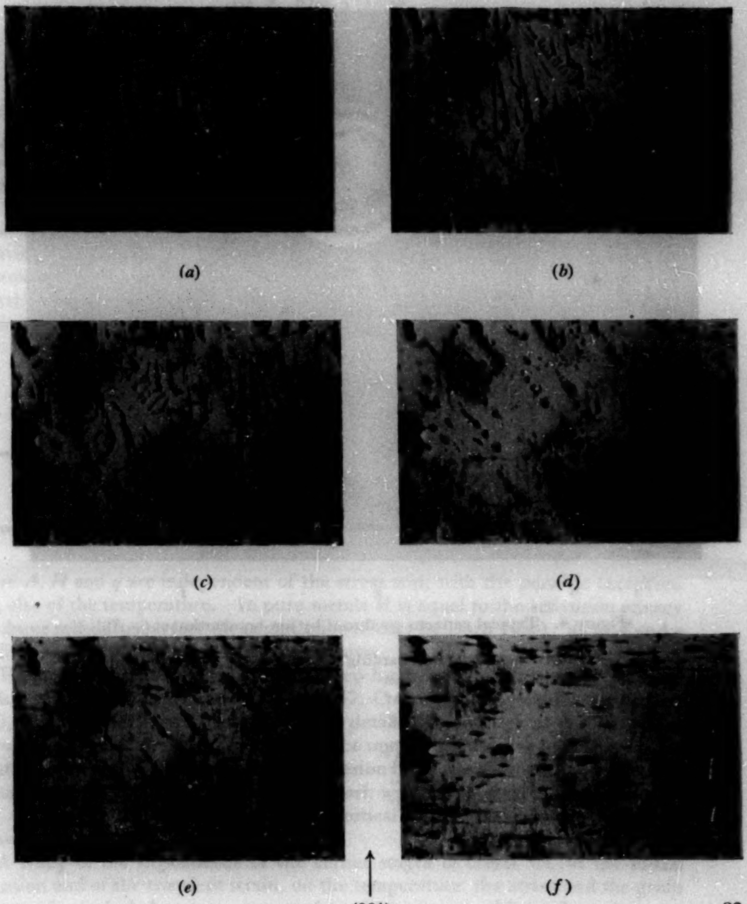


Figure 3. Appearance of domain patterns on (110) surface of silicon–iron as reverse field is increased from (*a*) to (*f*); crystal saturated at (*f*).

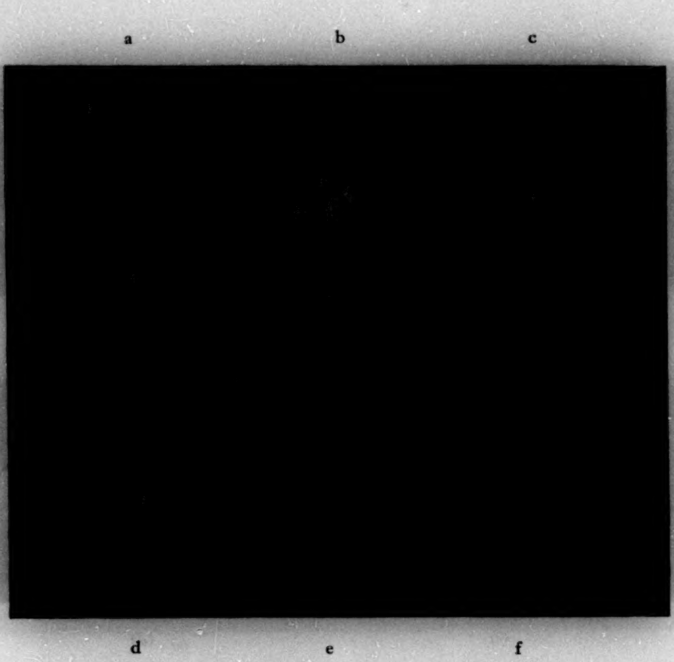


Figure 4. Typical patterns produced by ion bombardment (× 5).

Pattern	Target material	Bombarding ion	Target conditions kv	mA	hours
a	C	A^+	4·0	8·3	1·25
b	,,	Hg^+	3·2	1·9	25·7
c	,,	,,	3·0	1·8	5·3
d	Fe	A^+	3·0	1·0	3·1
e	C	,,	10·1	1·1	10·2
f	,,	,,	7·4	2·9	2·5

On the Mechanism of High-Temperature Creep in Metals with Special Reference to Polycrystalline Lead

By P. FELTHAM
The University, Leeds

MS. received 24th August 1955 and in revised form 22nd June 1956

Abstract. On the basis of Mott's suggestion that intracrystalline deformation in creep takes place by fine slip, and that the activation energy is that for the motion of screw dislocations impeded by jogs, expressions for the high-temperature transient and equilibrium creep rates are obtained in terms of the applied stress, the temperature, the grain size and the characteristics of the intragranular substructure. The equations give a good interpretation of the known creep behaviour of a number of metals as reported in the literature. The synthesis is based primarily on lead as, in conjunction with the present experiments, the available literaure provides a fairly comprehensive account of the creep of both single crystals and polycrystals of this metal.

§ 1. INTRODUCTION

THE formalism of the theory of rate processes adopted by Kauzmann (1941) for the interpretation of the dependence of the equilibrium creep rate $d\epsilon/dt$ in metals on the applied tensile stress σ and the temperature T yields the relation

$$d\epsilon/dt = A \exp(-H/\boldsymbol{k}T) \sinh(q\sigma/\boldsymbol{k}T),$$

where A, H and q are independent of the stress and, with the possible exception of q, also of the temperature. In pure metals H is equal to the activation energy of volume self-diffusion (see the compilation of data by Dorn 1956); as to A and q published experimental work is less conclusive.

The applicability of the equation to creep has been substantiated by work on numerous metals (Novick and Machlin 1947, Cottrell and Aytekin 1950, Fastov 1950, Feltham 1953, etc.), and if a more detailed knowledge of the nature of A and q were available, e.g. their dependence upon the grain size or intragranular substructures, it could provide a useful criterion for the assessment of the validity of current creep theories. The present work was designed to obtain such data and, if possible, to utilize them in a theoretical interpretation of, at least, one important mechanism of creep.

A study of the dependence of the earlier stages of creep, i.e. of the initial extension and of the transient strain, on the temperature, the stress and the grain size was also included as it was expected to yield relevant additional results.

Lead was chosen because several important aspects of the creep of single crystals and polycrystals of this metal had previously been considered by a number of workers, and because of the further convenience of working with a metal of low melting point.

§ 2. EXPERIMENTAL EQUIPMENT AND METHOD

2.1. *Material*

Batches of spectroscopically pure (99·99%) lead wires of 1·63 mm diameter, 12·4 cm long, were annealed *in vacuo* at 220°c for 12 hours, at 222°c for 24 hours, and at 310°c for 48 hours respectively; the mean grain diameters resulting from this treatment were $2L=0{\cdot}04$, 0·06 and 0·11 mm, the order being that of increasing annealing temperatures. Stabilization of the grain sizes at temperatures considerably above those eventually used in the experiments was intended to eliminate recrystallization and grain growth during creep, at least in so far as they were not induced by the creep deformation itself, and to reduce the intragranular strains to a minimum (Jenckel and Poeltsch 1938, Greenough and Smith 1950). Specimens with the largest grain size showed uniform equi-axed grains on sectioning; the dispersion of grain sizes in planar sections of specimens with the smaller grain sizes appeared to be somewhat greater. Presumably, nucleation and growth of grains were still appreciable towards the end of the heat treatment at the lower annealing temperatures.

The annealed wires were extremely soft and had to be handled with considerable care.

2.2. *Flow Tests*

The specimens were mounted in pin-vices (figure 1) to give an effective gauge length of 10·0 cm. Provided the pin-vices were finger-tight, slipping-out of the specimens did not occur, but additional support was, nevertheless, provided by attaching small spheres of solder to the specimen ends.

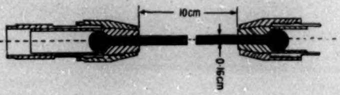


Figure 1. Mounting of the specimen in pin vices.

The specimens were enclosed in an electric oven thermostatically controlled to within ± 1°c; uniformity of temperature throughout the oven was ensured by means of a built-in fan rotating at low speed. Each specimen was annealed *in situ* for about 15 minutes at the flow temperature and the tensile load was then applied to it gradually in such a manner that 1–2 seconds elapsed between 'no load' and 'full load'. The stress was kept constant during the flow by means of an Andrade–Chalmers (1932) beam of the same design as described in earlier work (Feltham 1953). The specimen was connected to the beam by means of piano wire which passed through a small hole in the oven roof. The first scale reading was generally taken within 2–3 seconds after the full load had been applied. In order to determine the effect of temperature and applied stress on the strain rate three sets of experiments were carried out with each of the three groups of specimens of differing grain sizes. First, considering any given group, the temperature dependence of the tensile strain rate was investigated by allowing a number of specimens to flow at various temperatures, keeping the stress invariant. Similarly, by keeping the temperature invariant but changing the stress from specimen to specimen the stress dependence of the strain rate was found. Finally, to afford a check on the validity of any theoretical correlation of the results a few tests were made in each group with combinations of stresses and temperatures not provided for in the first two sets of experiments. Some typical flow curves are shown in figures 2 and 3. The strain is there defined as the extension, exclusive of the immediate initial extension, divided by the initial length of the specimen; the numbering of the curves corresponds to the classification of the results given in table 1. Since the experimentally determined transient strain, i.e. the strain following the instantaneous extension and preceding the stage of equilibrium flow, was generally less than 4–5%, conversion of the linear strain to the natural strain was not attempted in relation to this stage of flow, but in determining the equilibrium strain rates the natural strains were used.

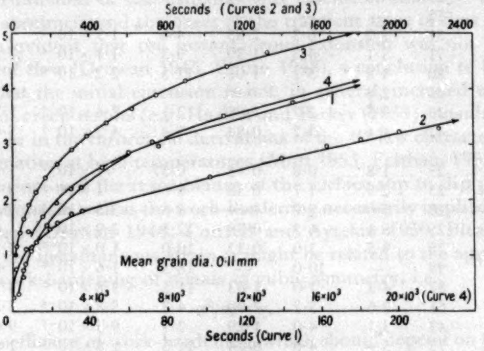


Figure 2. Creep curves of specimens with 0·11 mm grain diameter. (Numbers refer to column (1), table 1.)

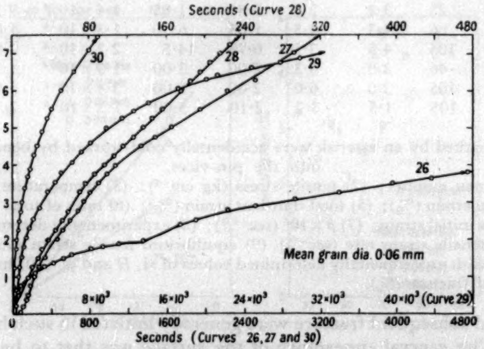


Figure 3. Creep curves of specimens with 0·06 mm grain diameter. (Numbers refer to column (1), table 1.)

§ 3. Experimental Results

3.1. *Microscopy*

Copious slip bands, which appeared to have formed predominantly in the earlier stages of the flow, were found on the surfaces of all specimens. Two or more slip systems were frequently observable, particularly near grain boundaries.

Table 1

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	100	25	3·7	3·1	0·84	7·5	6×10^{-5}	$5{\cdot}3\times10^{-5}$	11
2	74	25	1·8	1·7	0·94	2·5	8×10^{-7}	2×10^{-6}	5
3	87	25	2·7	4·1	1·52	4·3	$3{\cdot}5\times10^{-6}$	1×10^{-5}	>10·3
4	65	25	4·3	11	0·74	1·6	6×10^{-7}	$7{\cdot}5\times10^{-7}$	11
5	65	52	—	—	—	—	—	—	—
6	65	51	2·2	6·3	2·85	4·6	$1{\cdot}7\times10^{-5}$	$1{\cdot}1\times10^{-5}$	14
7	65	71	3·1	4·3	1·39	9·3	$1{\cdot}1\times10^{-4}$	$7{\cdot}2\times10^{-5}$	12
8	65	80	4·8	5·1	1·06	11·5	$2{\cdot}2\times10^{-4}$	$1{\cdot}9\times10^{-4}$	11
9	65	35	2·0	—	—	—	—	—	10
10	65	33	2·5	2·7	1·08	2·1	$2{\cdot}4\times10^{-6}$	$1{\cdot}7\times10^{-6}$	10
11	80	42	4·4	3·7	0·84	7·4	$4{\cdot}7\times10^{-5}$	$5{\cdot}6\times10^{-5}$	14
12	65	25	1·8	0·8	0·45	0·57	1×10^{-7}	$1{\cdot}3\times10^{-7}$	8
13	74	25	2·9	1·7	0·58	0·65	1×10^{-7}	$3{\cdot}7\times10^{-7}$	>8
14	100	25	5·0	5·7	1·14	3·6	$3{\cdot}3\times10^{-6}$	$9{\cdot}4\times10^{-6}$	13
15	115	25	8·1	6·3	0·78	12·5	$3{\cdot}9\times10^{-5}$	6×10^{-5}	13
16	125	25	9·5	3·0	0·32	10·0	$1{\cdot}0\times10^{-4}$	$5{\cdot}3\times10^{-5}$	14
17	65	99	—	10·0	—	18·5	$1{\cdot}6\times10^{-4}$	$1{\cdot}6\times10^{-4}$	16
18	65	82	4·3	7·0	1·63	11·5	$3{\cdot}8\times10^{-5}$	$3{\cdot}7\times10^{-5}$	16
19	65	61	2·6	5·2	2·00	4·3	$5{\cdot}8\times10^{-6}$	$5{\cdot}3\times10^{-6}$	11
20	65	43	3·1	4·0	1·29	2·7	$9{\cdot}5\times10^{-7}$	$9{\cdot}6\times10^{-7}$	13
21	85	73	5·4	4·0	0·74	11·5	$1{\cdot}8\times10^{-4}$	$1{\cdot}6\times10^{-4}$	12
22	65	25	3·3	5·2	1·58	0·80	$1{\cdot}4\times10^{-7}$	$1{\cdot}5\times10^{-7}$	>9
23	90	25	6·0	4·0	0·67	3·50	$4{\cdot}1\times10^{-6}$	$3{\cdot}2\times10^{-6}$	13
24*	100	25	6·1	5·1	0·84	3·50	$5{\cdot}7\times10^{-6}$	$1{\cdot}2\times10^{-5}$	11·5
25	125	25	8·1	4	0·50	15	$3{\cdot}0\times10^{-4}$	$2{\cdot}7\times10^{-4}$	11
26	80	25	3·2	2·6	0·81	1·85	$1{\cdot}4\times10^{-6}$	$9{\cdot}5\times10^{-7}$	11
27	65	66	4·3	4·5	1·05	6·50	$1{\cdot}3\times10^{-5}$	$1{\cdot}0\times10^{-5}$	15
28	65	105	4·5	3·5	0·78	14·5	$2{\cdot}3\times10^{-4}$	$2{\cdot}6\times10^{-4}$	16·5
29	65	46	3·0	6·3	2·00	3·00	$1{\cdot}3\times10^{-6}$	$1{\cdot}3\times10^{-6}$	>11
30	50	105	3·0	6·0	2·00	9·00	$3{\cdot}4\times10^{-5}$	$5{\cdot}6\times10^{-5}$	18
31*	40	105	1·5	3·2	2·10	5·00	$7{\cdot}3\times10^{-6}$	$2{\cdot}3\times10^{-5}$	>12

Specimens marked by an asterisk were accidentally cold-worked by bending on inserting into the pin-vices.

(1) specimen number; (2) tensile stress (kg cm^{-2}); (3) temperature (°c); (4) instantaneous initial strain (%); (5) total transient strain (%); (6) ratio of total transient strain to instantaneous initial strain; (7) $\beta\times10^{3}$ (sec$^{-1/3}$); (8) experimentally determined value of the equilibrium tensile strain rate (sec^{-1}); (9) equilibrium tensile strain rate calculated from equation (4) with experimentally determined values of A, H and q; (10) mean tensile strain at the time of fracture (%).

Necking and subsequent fracture were generally initiated in such highly deformed regions. The general appearance of the surface was that to be expected from a tensile test; slip bands varying in width from about 5 μ to well below the resolving power of the microscope were present. As the metal surface is unrepresentative of the mode of intragranular deformation, and as Hirst (1940, 1941) had previously studied slip occurring during creep in lead, no detailed measurements were made.

Measurements of the grain sizes of the specimens after flow disclosed that a diminution had taken place in some cases, particularly with specimens of the smaller grain sizes, which showed recrystallization during flow (e.g. curves 26, 27 in figure 3). The reduction occasionally approached 2 : 1; the initial nominal grain sizes were thus not preserved throughout the flow if creep-induced nucleation of new grains occurred.

3.2. *The Instantaneous Initial Extension*

In the present work an induction period was frequently found to intervene between the termination of the instantaneous extension observed on applying the load to the specimen, and the onset of the transient stage of flow (beta-flow). It was thus evident that the instantaneous extension was not part of the transient stage of flow (Orowan 1946, Smith 1948), a conclusion to be expected from the fact that the initial extension is not, in general, included either in the representation of creep results (e.g. Hazlett and Parker (1953) consider this point in some detail), or in the theoretical derivations of the $t^{1/3}$ law characteristic of the transient deformation at high temperatures (Mott 1953, Feltham 1955).

This conclusion, and the resemblance of the surface slip to slip produced by tensile deformations, as well as the work-hardening necessarily implied in recovery theories of creep (Orowan 1946, Cottrell and Aytekin 1950, Bleakney 1955), suggested that the instantaneous strain ϵ_I might be related to the applied stress σ by the law of work-hardening of metals of cubic symmetry, i.e.

$$\sigma^2/\epsilon_I = \chi, \qquad \ldots\ldots(1)$$

where χ is the coefficient of work-hardening, which should depend on the temperature and the grain size. Confirmation of this hypothesis is provided by the data in figure 4.

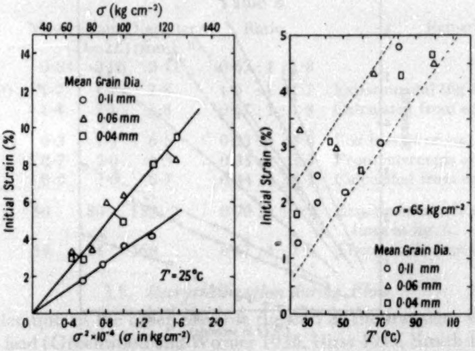


Figure 4. Dependence of the initial instantaneous strain on the stress at constant temperature, and on the temperature with the stress kept invariant.

3.3. *Induction Periods*

Induction periods of the type referred to above (e.g. curves 27 and 29 in figure 3) were also observed by Hirst in lead, by Rhines (1956) in aluminium, and by the author (unpublished results) in mild steel at about 1100°c. Existing evidence seems to show that 'geometric softening' (Cottrell and Aytekin 1950) takes place during this period, i.e. new slip planes are gradually brought into action by the strain rotations (and other modes of softening induced by recovery), mainly at grain boundaries.

3.4. *The Transient Creep*

In pure metals the transient strain can generally be expressed by the relation

$$\epsilon_{\mathrm{tr}} = \beta t^{1/3}, \qquad \ldots\ldots(2)$$

where, according to Mott (1953), and Feltham (1954, 1955)

$$\beta^3 = C\, d\epsilon/dt, \qquad \ldots\ldots(3)$$

where C is a constant and $d\epsilon/dt$ the equilibrium strain rate. If the latter is given by

$$d\epsilon/dt = A \exp(-H/\boldsymbol{k}T) \sinh(q\sigma/\boldsymbol{k}T), \qquad \ldots\ldots(4)$$

then we can write

$$\beta = \beta_0 \exp(-H_{\mathrm{tr}}/\boldsymbol{k}T) \sinh(q_{\mathrm{tr}}\sigma/\boldsymbol{k}T), \qquad \ldots\ldots(5)$$

where β_0 is a constant and, assuming $q_{\mathrm{tr}}\sigma \gg \boldsymbol{k}T$,

$$H_{\mathrm{tr}}/H = q_{\mathrm{tr}}/q = 1/3. \qquad \ldots\ldots(6)$$

That the functional form of equation (2) was obeyed is shown in figure 5. The zero of the time coordinate corresponds to the instant at which the initial extension was complete; pronounced induction periods tended to distort the curves, e.g. curve 4 of figure 5. Similar results were obtained with specimens of the smaller grain sizes, but on account of the distortion of the curves, due to the effects mentioned, the values of β obtained from them (table 1) are rather less reliable

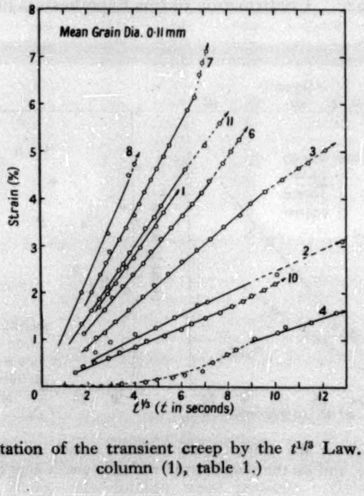


Figure 5. Representation of the transient creep by the $t^{1/3}$ Law. (Numbers refer to column (1), table 1.)

than those obtained from figure 5 for the specimens having the largest grain size. From the results shown in figures 6 and 7 one obtains $H_{\mathrm{tr}} = 7{\cdot}7\,\mathrm{kcal\,(g\,atom)^{-1}}$, and $q_{\mathrm{tr}} = 21{\cdot}5\,\mathrm{cal\,(g\,atom)^{-1}/kg\,cm^{-2}}$, both values independent of the grain size within the limits of the experimental error. With H and q as determined for the equilibrium flow (§ 3.5) the ratios (equation (6)) were both found to be equal to 0·29, which is close to the value of 1/3 expected from theory. The constant β_0 was determined separately for each group of specimens by means of equation (5) with the experimentally determined values of H_{tr} and q_{tr}, and conjugate values of β and σ, the latter two taken from the straight lines plotted in figures 6 and 7. The values of β_0 are given in table 2.

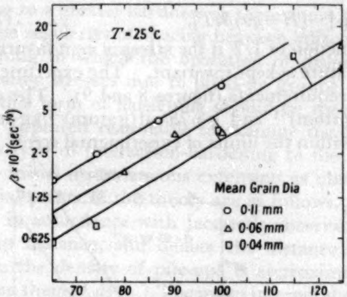


Figure 6. Dependence of β on the applied stress σ, and on the grain size.

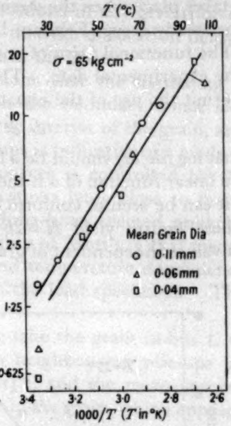


Figure 7. Dependence of β on the temperature and grain size.

Table 2

	Mean Grain Diameter ($D=2L$) (mm)			Ratio	Remarks
	0·04	0·06	0·11	0·67 : 1 : 1·8	
$\chi \times 10^{-5}$ (kg²)	1·7	1·7	2·8	1·0 : 1 : 1·7	Experimental (fig 4).
	1·4	2·1	3·8	0·67 : 1 : 1·8	Calculated from eqn (11).
$A \times 10^{9}$ (sec⁻¹)	0·3	1·3	6·5	0·23 : 1 : 5·0	For best fit of eqn (4).
	0·7	2·0	6·5	0·35 : 1 : 3·3	From intercepts of lines in fig. 8.
	0·9	2·0	6·7	0·44 : 1 : 3·3	Calculated from eqn (22).
β_0 (sec⁻¹)	56	80	153	0·70 : 1 : 1·9	Experimental; from intercepts of lines in fig. 6.
	59	88	158	0·67 : 1 : 1·8	Theoretical; eqn (25).

3.5. *Recrystallization during Flow*

An acceleration of the creep towards the end of the transient stage has been observed in lead (Greenwood and Worner 1936, Hirst 1941, Smith 1941, Andrade 1948) and in other metals (Feltham 1953, Grant and Bucklin 1953). Andrade (1948) established this to be due to recrystallization. Such effects were found to occur in some of the specimens (e.g. curves 26, 27, 28 of figure 3) of 0·04 and 0·06 mm nominal grain size and were probably induced by high stress concentrations at grain boundaries arising out of the progressive hardening in the early stages of creep. Recrystallization is not invariably present in creep, and no fundamental significance can be ascribed to it.

3.6. *The Equilibrium Flow*

The mode of flow following the transient stage, or the stage of recrystallization if such takes place, is characterized by the constancy of the strain rate under conditions of contant stress, and by the implied non-occurrence of progressive work-hardening. The transition from the transient to the steady stage is smooth, and takes place when the strain rate in the former reaches the value characteristic of the steady creep (Feltham 1954).

The functional form of equation (4) was found to provide a good correlation of the experimental data. The tensile stresses were in all cases sufficiently high to permit the use of the equation in the form

$$d\epsilon/dt=\tfrac{1}{2}A\exp\left[-(H-q\sigma)/\boldsymbol{k}T\right], \qquad \ldots\ldots(7)$$

so that $\log(d\epsilon/dt)$ should be a linear function of $1/T$ if the stress is kept invariant, and a linear function of σ if the temperature is kept invariant. The experimental result can be seen to conform to both requirements (figures 8 and 9). The data in these figures yield $H=26{\cdot}4$ kcal (g atom)$^{-1}$ and $q=75$ cal (g atom)$^{-1}$/kg cm^{-2}, both values independent of grain size within the limits of experimental error.

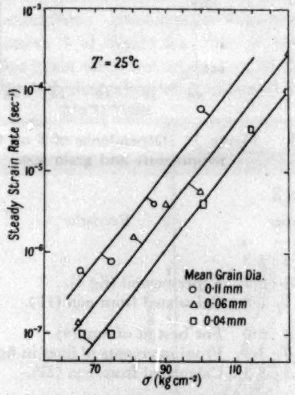


Figure 8. The stress dependence of the equilibrium strain rate at a constant temperature (25°c).

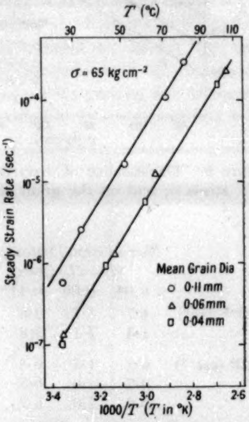


Figure 9. The temperature dependence of the equilibrium strain rate with the stress kept invariant at 65 kg cm^{-2}.

§ 4. The Creep Mechanism

4.1. *The Instantaneous Initial Extension*

The interpretation of the hardening which takes place on applying the load to the specimen (figure 4) will here be based on the experimental observations of

Jacquet (1954 a, b) on the behaviour of dislocations in alpha-brass subjected to small tensile strains, and on the theoretical treatment of the associated mode of hardening developed by Mott (1952, 1953).

The features in Jacquet's micrographs which are most interesting from the point of view of the present work are: (*a*) Arrays of dislocations, which appear to originate predominantly in the central regions of the grains (the latter being of about the same size as in the lead used in this research), traverse the grain and pile up at the boundaries on the slip-planes containing their respective sources; (*b*) the pile-ups are approximately equidistant, the spacing between adjacent ones being a few microns; (*c*) the number of dislocations in any given pile-up is of order 10–100; (*d*) a large fraction of the dislocations in any given pile-up lies close to the boundary at which it is held up, so that the dislocation density is much greater in the boundary zones than elsewhere in the grains.

It is possible to conclude from these observations that the hardening is heterogeneous, the greater density of dislocation in the boundary zones also leading to a greater hardness in these zones than in the interior of the grain, and that the regularity of spacing between adjacent pile-ups is indicative of a mode of hardening in which the operation of dislocation sources is controlled by the long-range stresses due to piled-up groups of dislocations lying close to them. That this form of hardening should be specific to alpha-brass seemed unlikely, and it appeared reasonable to examine the applicability of Mott's (1953) theory of this type of interaction-hardening to the stress and temperature dependence of the initial instantaneous extension as observed in the lead specimens. The essential points of the theory are as follows.

If, in accordance with Jacquet's observations, we take the grain radius L as the slip distance, and denote the distance between neighbouring pile-ups by d, then the density of pile-ups is approximately $(Ld)^{-1}$, and the mean spacing between them $(Ld)^{1/2}$. The mean internal shear stress in the grain, which opposes the applied shear stress τ, is then

$$\tau_1 = -\frac{Gbn}{2\pi(Ld)^{1/2}} \qquad \ldots\ldots(8)$$

where n is the number of dislocations per pile-up, G is the shear modulus, and b the Burgers vector. The negative sign shows τ_1 to oppose τ, the sense of the latter having been assumed positive.

Now the plastic shear strain is

$$\gamma \simeq nb/d, \qquad \ldots\ldots(9)$$

and on writing $\tau_1 = -\tau$ (which is permissible except for very small stress, viz. Mott 1953), and putting

$$\tau \simeq \tfrac{1}{2}\sigma \quad \text{and} \quad \epsilon_1 \simeq \gamma, \qquad \ldots\ldots(10)$$

(observing that the strain as defined by equation (9) is the 'engineering' strain), where σ and ϵ_1 are the tensile stress and strain respectively, we obtain the parabolic work-hardening law (equation (1)) with the coefficient of work-hardening given by

$$\chi = G^2bn/\pi^2L. \qquad \ldots\ldots(11)$$

In this theory the dislocations are assumed to be generated dynamically. They are stopped by the boundary when

$$n = \tau_Y(2\pi L/Gb), \qquad \ldots\ldots(12)$$

where τ_Y is the yield stress in shear of the undeformed metal. The number of dislocations per pile-up does not therefore depend upon the applied stress. Taking room temperature values, $\tau_Y = 10\,\mathrm{kg\,cm^{-2}}$ (Feltham 1956), $L = 3 \times 10^{-3}\,\mathrm{cm}$, $G = 5 \times 10^4\,\mathrm{kg\,cm^{-2}}$, $b = 3{\cdot}5 \times 10^{-8}\,\mathrm{cm}$, we find $n \simeq 100$. As the lengths of the pile-ups are considerably less than the size of the grains if the latter are as large as those in the lead specimens, n would not be expected to depend significantly on L, which would imply (equation (12)) that τ_Y should be approximately inversely proportional to L. Since also (from equations (11) and (12))

$$\chi = (2/\pi) G \tau_Y \qquad \ldots\ldots(13)$$

the coefficient of work-hardening should be smaller for the specimens with the larger grains than for the others. Figure 4 shows this to be the case. Its temperature dependence should also approximate to that of the yield stress τ_Y. Reliable data on the temperature dependence of the yield stress of polycrystalline lead were not available to the author, but an estimate of the magnitude of the effect could be obtained from Neurath's and Koehler's (1951) measurement of the yield stress of single crystals of lead. This they found to be approximately halved on increasing the temperature from 25°c to 100°c, which is in reasonable agreement with the variation of χ shown in figure 4, particularly if it is remembered that G will also decrease as the temperature is raised.

The spacing between adjacent pile-ups (i.e. active slip planes) at a tensile stress of $65\,\mathrm{kg\,cm^{-2}}$ then works out to $d \simeq 1\,\mu$. Intersection of slip systems would thus lead to the formation of 'mosaic' blocks with linear dimensions of about one micron. In polycrystalline lead their existence was established by Hirst (1941) in creep, and by Kolontsova (1953) as resulting from small tensile deformations.

4.2. *The Activation Energy of the Equilibrium Flow*

In a recent critical analysis of creep mechanisms Mott (1956) concludes that none could be considered adequate, and outlined a new mechanism of high temperature equilibrium creep in which the motion of screw dislocations containing jogs is rate determining. By means of this model he was able to account satisfactorily for the commonly observed magnitudes of the parameters A, H and q appearing in equation (4). If τ is the applied shear stress acting on a screw dislocation in which jogs occur at an average distance l apart and b is the Burgers vector, then the activation energy for the movement of such a dislocation is

$$Q(\sigma) = H - \tfrac{1}{2}\sigma b^2 l, \qquad \ldots\ldots(14)$$

on writing $\frac{1}{2}\sigma = \tau$. H is the activation energy for volume self-diffusion, and appears in equation (14) because non-conservative motion of jogs due to the intersection of two screw dislocations necessitates the formation and dispersion of vacancies. On comparing the stress dependent terms of equations (7) and (14) we find

$$q = \tfrac{1}{2} l b^2 \qquad \ldots\ldots(15)$$

which, on substituting the measured value of

$$q (= 75\,\mathrm{cal\,(g\,atom)^{-1}/kg\,cm^{-2}} = 5{\cdot}2 \times 10^{-21}\,\mathrm{cm^3})$$

gives $l \simeq 0{\cdot}1\,\mu$, i.e. about one order of magnitude below the block dimensions, which is a reasonable value.

The activation energy for volume self-diffusion in lead is $27\,\mathrm{kcal\,(g\,atom)^{-1}}$ according to Novick (1951), and $25{\cdot}7\,\mathrm{kcal\,(g\,atom)^{-1}}$ according to recent measurements made by Okkerse (1954). The mean value of the two is in very good agreement with the value of $H(=26{\cdot}4\,\mathrm{kcal}\,(\mathrm{g\,atom})^{-1})$ obtained in the present work (§ 3.6). (Buffington's and Cohen's (1952) observation that self-diffusion in alpha-iron is accelerated if the metal is subject to creep could readily be explained if, as is assumed to be the case, vacancies are generated in the creep process.)

4.3. *The Role of Recovery and Strain Hardening*

Cottrell and Aytekin (1950) have shown convincingly that high temperature creep must be regarded as resulting from simultaneous hardening and recovery under stress, so that according to whether the creep rate is decreasing, constant, or increasing, the rate of hardening is greater, equal, or less than the rate of recovery. The condition for equilibrium creep is then (using the notation of equations (8) and (10)):

$$-d\tau_1=d\tau=\frac{\partial\tau}{\partial t}\,dt+\frac{\partial\tau}{\partial\gamma}\,d\gamma=0, \qquad \ldots\ldots(16)$$

which defines the shear rate $d\gamma/dt$, and hence the tensile strain rate (equation (10)). In polycrystalline lead subject to creep at 22°c under a constant stress of 112 kg cm^{-2} Kennedy (1955) found the activation energies of equilibrium creep and recovery to be identical, as would be expected to be the case on the basis of Cottrell's and Aytekin's generalization.

Dealing first with the rate of recovery $\partial\tau/\partial t$ the assumption will be made, in accordance with well-known experimental evidence of the enhanced activity in grain boundary zones during creep, that recovery occurs foremost at grain boundaries which are sinks for dislocations and vacancies. Now it can be shown by a method used by McLean (1956) that under conditions of steady creep at stresses as used in the present work a dislocation could penetrate a boundary only in special cases, for instance, tilt boundaries with a 'tilt' angle of less than approximately 3°. Such a mode of loss of dislocations from grains would be unimportant in pratice. In any case, as it would result in the transfer of a dislocation to another grain, rather than the annihilation of the dislocation, recovery would not, in effect, take place except if the dislocation passed out of the metal at the specimen surface. Since we assume the rate of generation of dislocations to be equal to the rate of loss in equilibrium flow, such a mechanism would imply a marked dependence of the creep rate on the shape and dimensions of the specimen. This is not, in general, observed. Similarly, McLean's (1952) proposal, that the continual accumulation of dislocations in subgrains or polygon walls could be regarded as the recovery mechanism whereby a 'removal' of dislocations occurred in aluminium during creep, can be considered valid at best in the stage of transient creep, for it implies progressive hardening, which is irreconcileable with the steady-state condition (equation (16)). We conclude that grain boundary migration, recrystallization, etc., in so far as they are not integral parts of processes resulting in the loss of dislocations, cannot determine the equilibrium creep rate, and assume that recovery takes place essentially by cross-slip in such a way that dislocations of opposite signs mutually annihilate. Such a process is not inhibited by geometrical factors such as the magnitude of the tilt between adjacent grains, as the stress between attracting dislocations can reach the theoretical shear strength of the metal.

An example of the dissolution of dislocation loops initiated by cross-slip of the screw components is shown in figure 10. Two dislocation loops of opposite signs, each belonging to a group originating from a common source, lie in parallel slip planes separated by the distance AA′. The screw components AD, BC and A′D′, B′C′ can move in the planes AA′D′D and BB′C′C and will coalesce as shown for one set AD and A′D′, annihilating over the length PQ. P and Q then move into alignment with AA′ and DD′ respectively. BC and B′C′ are similarly annihilated. (Jogs are not shown in figure 10). We are therefore left with two rectangular edge dislocations AA′B′B and CC′D′D, the movement of which is not impeded by jogs. If AB is positive, i.e. the plane of extra atoms lies above AB, as indicated in the figure, then the movement of AA′B′B (by 'pencil glide', Nye 1949) into the adjacent grain boundary would result in removing a sheet of atoms of area ABB′A′ from the grain surface. The process can thus lead to the accumulation of vacancies near grain boundaries (Greenwood *et al.* 1954, Feltham 1953) or, in the case of negative dislocations, to the accumulation of interstitial atoms with the associated increased local stresses. The latter may then be reduced by recrystallization, cross-slip across the boundary, etc. The initiation of the recovery process by screw dislocations containing jogs would require an activation energy as given by equation (14), since the initial intensity of interaction between dislocations is characterized by the internal stress which, as will be shown below, is of the same order as the applied shear stress τ. On account of the closer approach of the dislocations at later stages of the process the activation energy is then reduced, so that the rate of approach increases. Hence the rate of initiation, which is the slowest process, determines the rate at which dislocations are lost from these groups. If n^* dislocations belong to each group, the rate is defined by dn^*/dt. If cross-slip of this type were initiated simultaneously in m of the n^* dislocations, and if each dislocation contained only one jog, then dn^*/dt would be given by $m(\boldsymbol{k}T/\boldsymbol{h})\exp[-(H-q\sigma)/\boldsymbol{k}T]$. If, as must be assumed in general, each screw dislocation contains several jogs which are not simultaneously activated, we write

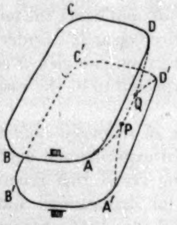


Figure 10. Mutual annihilation of the screw components of the dislocation loops ABCD and A′B′C′D′ by cross slip.

$$dn^*/dt=\alpha(\boldsymbol{k}T/\boldsymbol{h})\exp[-(H-q\sigma)/\boldsymbol{k}T], \qquad \ldots\ldots(17)$$

where $\alpha=m/s$, s being of the same order as the number of jogs per screw dislocation. A reasonable estimate would be $0{\cdot}1<\alpha<10$. We now have to consider the hardening during the equilibrium flow, We assume that about half of the dislocations in the pile-ups formed on applying the load are dispersed during the stage of the transient flow (Stroh 1953), which leads to a more uniform distribution of dislocations throughout the grains and to a levelling of the internal stress peaks. Such a distribution provides the prerequisite to fine slip (i.e. slip too fine to be observable by ordinary optical methods). The theory of this type of slip (Mott 1953) will therefore be used to obtain $\partial\tau/\partial\gamma$ (equation (16)).

4.4. *The Equation of the Equilibrium Flow*

We begin by considering a grain in the form of a cube of side $D=2L$ in which the density of active slip sources is N. The (internal) shear strain due the operation of the ND^3 sources is then

$$\gamma_{\text{int}}=-(ND^3)(bn^*/D)=-ND^2bn^*. \qquad \ldots\ldots(18)$$

The distance between adjacent groups is of order D/ND^3, and the mean of this and D, namely $(ND)^{-1/2}$, if taken as the mean distance between any given group and all others, gives for the internal stress the relation

$$\tau_{\text{int}}\simeq-\tau\simeq Gbn^*/2\pi(ND)^{-1/2}. \qquad \ldots\ldots(19)$$

Noting that if n^* is reduced by the escape of dislocations the internal strain is reduced but a plastic shear due to slip is introduced such that

$$d\gamma=-d\gamma_{\text{int}}, \qquad \ldots\ldots(20)$$

we obtain from equations (17), (18) and (19) on substitution into equation (16) (using the 'sinh' to replace '$\frac{1}{2}$ exp' to allow for small values of $q\sigma/\boldsymbol{k}T$):

$$d\epsilon/dt=A\exp(-H/\boldsymbol{k}T)\sinh(q\sigma/\boldsymbol{k}T) \qquad \ldots\ldots(21)$$

with

$$A=2\alpha ND^2b(\boldsymbol{k}T/\boldsymbol{h}). \qquad \ldots\ldots(22)$$

An estimate of N can be obtained by considering a cubical grain of side $D=10^{-2}$ cm, with one active slip system, as in deriving equation (18). If the slip planes are about 1 μ apart, each having originated at one source, then the crystal contains 100 sources, i.e. for a unit cube we have $N=10^8$. This estimate is identical with that made by Mott (1953) in dealing with fine slip. With this value of N, taking $\boldsymbol{k}T/\boldsymbol{h}=6\cdot3\times10^{12}$ (i.e. $T=300^\circ$K), $b=3\cdot5\times10^{-8}$, and $\alpha=1\cdot25$, the experimentally determined and the theoretically derived values of A (equation (21)) are in good agreement (table 2).

4.5. *The Transient Flow*

There is a considerable amount of evidence in support of Mott's (1953) derivation of the equation of the transient flow at high temperatures (equations (2) and (3)) (e.g. Feltham 1954), and his method is here adopted with a difference however, in the interpretation of the details of the mechanism.

In terms of the parameters previously defined, Mott's analysis yields

$$\beta^3=(3ND^2b)(\tau_1/\theta)^2dn^*/dt, \qquad \ldots\ldots(23)$$

where θ is the coefficient of linear work hardening $\partial\tau/\partial\gamma$ (obtainable from equations (18)–(20) and equal to $G/2\pi N^{1/2}L^{3/2}$), and $\tau_1=Gb/l_1$, where $l_1\simeq L/n^*$.

Owing to the partial dispersion of the dislocations in the pile-ups (§ 4.3) n^* will be rather less than n (equation (12)). The fact that the ratio of the total transient strain to the initial strain due to the instantaneous extension $\epsilon_{\text{tr}}/\epsilon_{\text{i}}$ is of order unity (table 1) indicates that up to one half of the dislocations may be lost in this stage of the flow. Hence we shall take $n^*=60$ in the subsequent calculation. From equations (17), (22) and (23):

$$\beta^3=3\left(\frac{\tau_1}{\theta}\right)^2\frac{d\epsilon}{dt}=\left[3\left(\frac{\tau_1}{\theta}\right)^2A\right]\exp\left(-\frac{H}{\boldsymbol{k}T}\right)\sinh\frac{q\sigma}{\boldsymbol{k}T}. \qquad \ldots\ldots(24)$$

As far as the activation energy and its stress-dependent part is concerned the agreement between experimental values and those predicted by equation (24) has been considered in § 3.4, and was found to be good. It therefore remains only to consider the factor β_0, where $\beta_0{}^3=3(\tau_1/\theta)^2A$. On writing $(\tau_1/\theta)^2=4\pi^2b^2NL(n^*)^2$, obtained from equation (23) *et seq.* we have

$$\beta_0{}^3=12\pi^2b^2NLA(n^*)^2. \qquad \ldots\ldots(25)$$

On substituting into equation (25) the theoretically determined values of A (table 2), and n^* and N as given above (i.e. 60 and 10^8 cm^{-3} respectively), very good agreement between the theoretically determined and observed values of β_0 is obtained (table 2).

§ 5. DISCUSSION

The magnitude of A (equation (22)) obtained with various pure metals in high temperature creep (above about $\frac{1}{2}T_{\text{melt}}$°K) varies over the range 10^5–10^{10} sec^{-1}. Considerable variability is to be expected in view of its functional form; it is also liable to considerable error due to comparatively small errors in the estimation of H and q. In fact, from the point of view of assessing creep theories the parameter q is of greater significance than A.

As to the stress dependence of q, data obtained with stresses covering a 4 : 1 range (e.g. Tyte 1938, 1939, McVetty 1943) show conclusively that, on interpreting the equilibrium creep rate by means of equation (21), q is independent of the applied stress.

In relation to its temperature dependence, Smith's (1941) data on the creep of lead (99·999%) obtained over a temperature range exceeding and including the one used in the present work, suggested to the present author that q could be regarded as essentially temperature independent. Such an assumption was found to lead to satisfactory results also in earlier work on the creep of gamma-iron in the range $0{\cdot}70<T/T_{\text{melt}}<0{\cdot}87$, with q measured in approximately the middle of the range (Feltham 1953). However, it is clearly apparent from the work of Tyte (1938, 1939) on lead and tin, and from creep results obtained with aluminium and other metals by Kauzmann (1941) as well as by Dorn (1956) that the temperature dependence of q cannot, in general, be neglected. This was also surmised by Mott (1956). The observed temperature dependence of q invariably shows an increase of q with temperature (Novick and Machlin 1947, Fastov 1950). Kauzmann suggested the relation $q=q_0\exp(gT)$, where q_0 and g were characteristic of the particular metal considered. Over narrow ranges of the temperature the relation $q=q_0{}'T$ may be applicable in some cases, thus leading to an apparent temperature independence of the stress-dependent part of the activation energy (Dorn 1956). The data in table 3, as well as the parallelism between the temperature dependence of q in zinc and the subgrain dimensions in aluminium subject to creep at a constant strain rate (figure 11) suggest a dependence of q on the dimensions of intragranular substructure which, in the equilibrium flow, is determined primarily by the temperature. (Some evidence on this point has been obtained by Dorn (1956).) A detailed study of the dependence of q on temperature, crystal structure, and on the presence of other elements and phases would be particularly valuable.

Table 3

	q	T/T_m	Reference
Al (99·99%)	80	0·57	Dorn 1956
Al-alloy RR59 (precip. hardened)	5	0·46†	Johnson & Frost 1952
AgBr (single crystal)	94	0·86†	From data of Christy 1954
AgBr	7	0·86†	From data of Christy 1954
Cu+3% Ag (precip. hardened)	3	0·40†	From data of Olds 1954
Pb (99·999%)	60	0·55†	From data of Smith 1941
Pb (99·99%)	75	0·56†	This paper.
Sn (99·9%)	45	0·58	From data of Tyte 1938
Zn (99·99%)	64	0·55†	Cottrell and Aytekin 1950
Zn (99·99%) (single crystal)	2500	0·55†	Cottrell and Aytekin 1950
Zn (99·999%) (single crystal)	20500	0·43	From data of Read and Tyndall 1946

† Denotes the mean temperature if a range of temperatures is covered. q in cal (g atom)$^{-1}$/kg cm^{-2}; T, T_m in °K.

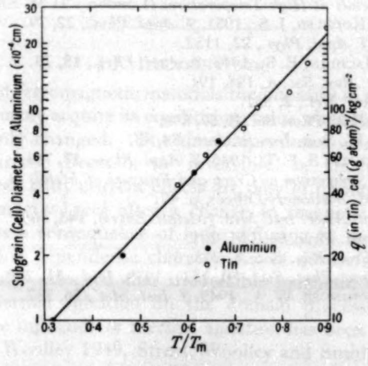


Figure 11. The dependence of the cell dimensions in aluminium, and of q in tin on the absolute temperature, the melting point T_m in each case referring to the respective metal. (Al, Wood and Rachinger 1949, Sn, Tyte 1938.)

ACKNOWLEDGMENTS

The author wishes to thank Professor N. F. Mott for valuable comments while the paper was being written.

REFERENCES

ANDRADE, E. N. DA C., 1948, *Nature, Lond.*, **162**, 410.
ANDRADE, E. N. DA C., and CHALMERS, B., 1932, *Proc. Roy. Soc.* A, **138**, 348.
BLEAKNEY, H. H., 1955, *Canad. J. Tech.*, **33**, 56.
BUFFINGTON, F. S., and COHEN, M., 1952, *Trans. Amer. Inst. Min. (Metall.) Engrs*, **194**, 859.
CHRISTY, W., 1954, *Acta. Met.*, **2**, 284.
COTTRELL, A. H., and AYTEKIN, V., 1950, *J. Inst. Met.*, **77**, 389.
DORN, J. E., 1956, ***Symposium on Creep and Fracture of Metals at High Temperatures*** (London : H.M. Stationery Office), p. 89.

Fastov, N. S., 1950, *Zhur. Tekh. Exp. Fiz.*, **20,** 543.
Feltham, P., 1953, *Proc. Phys. Soc.* B, **66,** 865; 1954, *Phil. Mag.*, **45,** 9; 1955, *Brit. J. Appl. Phys.*, **6,** 26; 1956, *Nature, Lond.*, **177,** 632.
Grant, N. J., and Bucklin, A. G., 1953, *Trans. Amer. Soc. Metals,* **35,** 151.
Greenough, G. B., and Smith, E. M., 1950, *J. Inst. Met.*, **77,** 435.
Greenwood, J. N., Miller, D. R., and Suiter, J. W., 1954, *Acta Met.*, **2,** 250.
Greenwood, J. N., and Worner, H. K., 1936, *J. Inst. Met.*, **64,** 135.
Hazlett, T. H., and Parker, E. R., 1953, *J. Met.*, **6,** 318.
Hirst, H., 1940, *Proc. Austral. Inst. Min. Met.*, **118,** 101, **120,** 11, 77; 1941, *Ibid.*, **121,** 11, **121,** 29.
Jacquet, P. A., 1954 a, *Acta Met.*, **2,** 752; 1954 b, *Ibid.*, **2,** 770.
Jenckel, E., and Poeltsch, H., 1938, *Z. Metallkde*, **30,** 396.
Johnson, A. E., and Frost, N. E., 1952, *J. Inst. Met.*, **81,** 93.
Kauzmann, W., 1941, *Trans. Amer. Inst. Min. (Metall.) Engrs*, **143,** 57.
Kennedy, A., 1955, *Proc. Phys. Soc.* B, **68,** 257
Kolontsova, E. V., 1953, *Dok. Akad. Nauk, SSSR*, **93,** 795.
McLean, D., 1952, *J. Inst. Met.*, **81,** 287; 1956, *Symposium on Creep and Fracture of Metals at High Temperatures* (London : H.M. Stationery Office), p. 73.
McVetty, P. G., 1943, *Trans. Amer. Soc. Mech. Engrs*, **65,** 761.
Mott, N. F., 1952, *Phil. Mag.*, **43,** 1115; 1953, *Ibid.*; **44,** 742; 1956, *Symposium on Creep and Fracture of Metals at High Temperatures* (London : H.M. Stationery Office), p. 21.
Neurath, P. W., and Koehler, J. S., 1951, *J. Appl. Phys.*, **22,** 79.
Novick, A. S., 1951, *J. Appl. Phys.*, **22,** 1182.
Novick, A. S., and Machlin, E. S., 1947, *J. Appl. Phys.*, **18,** 79.
Nye, J. F., 1949, *Proc. Roy. Soc.* A, **198,** 191.
Okkerse, B., 1954, *Acta Met.*, **2,** 551.
Olds, G. C. E., 1954, *Proc. Phys. Soc.* B, **67,** 840.
Orowan, E., 1946, *J. W. Scot. Iron St. Inst.*, **54,** 45.
Read, T. A., and Tyndall, S. P. T., 1946, *J. Appl. Phys.*, **17,** 713.
Rhines, F. N., 1956, *Symposium on Creep and Fracture of Metals at High Temperatures* (London : H.M. Stationery Office), p. 47.
Smith, A., 1941, *Trans. Amer. Inst. Min. (Metall. Engrs)*, **143,** 165.
Smith, C. L., 1948, *Proc. Phys. Soc.*, **61,** 201.
Stroh, A. N., 1953, *Proc. Roy. Soc.* A, **218,** 391.
Tyte, L. C., 1938, *Proc. Phys. Soc.*, **50,** 153; 1939, *Ibid.*, **51,** 203.
Wood, W. A., and Rachinger, W. A., 1949, *J. Inst. Met.*, **76,** 237.

A Comparison of Magnetic Viscosity in Isotropic and Anisotropic High Coercivity Alloys

By R. STREET† and J. C. WOOLLEY‡

† University of Sheffield. ‡ University of Nottingham.

MS. received 1st *June* 1956

Abstract. Measurements of magnetic viscosity in equiaxed isotropic and anisotropic Alcomax III specimens have been made; the anisotropy was induced by cooling the material, from the single phase region, in a magnetic field of 4000 Oe. The results are discussed in terms of the concept of the thermal activation of irreversible domain boundary wall movements. Comparisons of the results of measurements on anisotropic and isotropic specimens are in reasonable agreement with theoretical predictions.

§ 1. INTRODUCTION

For a number of ferromagnetic materials the intensity of magnetization does not instantaneously acquire its equilibrium value when the applied magnetic field is suddenly changed. The time dependence of magnetization may be due to eddy current effects or to 'reversible' or 'irreversible' magnetic viscosity. In general, eddy current effects are only of importance for relatively short periods of time (<1 sec) after a change in magnetic field. Reversible magnetic viscosity is a consequence of ionic diffusion in the crystalline lattice and has a temperature dependence characteristic of diffusion processes (Snoek 1947). It is generally accepted that irreversible magnetic viscosity is due to the influence of thermal agitation on the domain processes responsible for magnetization. The influence of thermal agitation has been considered by the authors (Street and Woolley 1949, Street, Woolley and Smith 1952) in terms of the energy required to activate irreversible domain processes, and it follows that the time dependence of the intensity of magnetization is given by

$$I(t) = S \ln t + \text{const.} \qquad \ldots\ldots(1)$$

S is a constant, the value of which depends on the material of the specimen, the point of measurement on the hysteresis cycle, the temperature of measurement and the demagnetizing factor of the specimen; t is the time interval which has elapsed after the sudden change in applied field. The results of experiments on irreversible magnetic viscosity are in good agreement with equation (1). An alternative approach has been given by Néel (1950, 1951) in which it is assumed *a priori* that the effect of thermal agitation may be represented by a 'fictitious' field $H_f(t)$ which is then shown to be a linear function of $\ln t$; the associated variation of intensity of magnetization is then

$$I(t) = \chi_{\text{irr}} H_f(t) + \text{const.} \qquad \ldots\ldots(2)$$

where χ_{irr} is the irreversible differential susceptibility at the point of measurement on the hysteresis cycle. The formal connection between the two approaches is derived from equations (1) and (2) from which $H_{\mathrm{f}}(t)=(S/\chi_{\mathrm{irr}})\ln t$. It can be shown that

$$\frac{S}{\chi_{\mathrm{irr}}}=\frac{kT}{q} \qquad \ldots\ldots(3)$$

where T is the absolute temperature of the specimen and q is a constant,† related to the volume of material involved in a single activation process. For example, if the specimen consists of a collection of identical non-interacting single domain ellipsoidal particles aligned with the major axes along the direction of the field, $q=vJ_{\mathrm{s}}$ where v is the volume of each ellipsoid and J_{s} is the spontaneous magnetization appropriate to the temperature of measurement. Experimental observation shows that for certain magnetic materials, in particular Alnico, q is largely independent of the point of measurement on the hysteresis cycle and of the temperature over a wide range (Street, Woolley and Smith 1952).

Magnetometric methods of measurement have usually been employed in investigations of magnetic viscosity, imposing limitations on the physical form of the specimens and on the maximum applied magnetic field, as the latter must necessarily be derived from solenoids. The present communication is concerned with magnetic viscosity in isotropic and anisotropic Alcomax III, specimens of which may be conveniently cast into block form but not into the long thin cylinders suitable for magnetometric investigation. In addition the coercivity of such specimens may rise to 800 Oe necessitating saturating fields of 4000 Oe or greater. It was therefore necessary to develop new techniques of measurement of magnetic viscosity. The first part of this paper is concerned with descriptions and the testing of these techniques.

§ 2. METHOD AND APPARATUS

The principle of the method is to measure the time variation of the e.m.f. $E(t)$ induced in a coil wound round the specimen, held between the poles of an electromagnet, with a steady magnetic field H applied to the specimen. From equation (1)

$$E(t)=-4\pi NA\frac{dI(t)}{dt}=-\frac{4\pi NAS}{t} \qquad \ldots\ldots(4)$$

where N is the number of turns on the coil, A the cross-sectional area of the specimen, t is the time which has elapsed since the magnetic field H was established and the value of S depends *inter alia* on the effective demagnetization coefficient of the specimen; a measure of this is the drop in magnetic potential across the contacts between the faces of the specimen and the pole faces of the electromagnet. If $E(t)$ is sufficiently large it may be measured directly on a recording potentiometer, or alternatively $E(t)$ may be amplified by means of a galvanometer amplifier (see Simpson and Tredgold 1953). The magnetic potential drops at the junctions were effectively reduced to zero by careful machining of the faces of the specimens and the pole tips. The condition for zero magnetic potential drop at the junctions was tested by the method adopted for a similar purpose in precision permeameters (see for example Mee and Street 1954). The specimens were thus operated under closed magnetic circuit conditions, with zero demagnetization factor, and the constant S of equation (4) is written S_0.

---

† q is defined by the relation $\Delta E=-q\Delta H$ where ΔE is the average change of activation energy of the domain processes consequent upon an increase ΔH in applied field.

Equation (4) is only applicable to an idealized situation in which it is assumed that instantaneous changes of applied field are possible. The origin of the time t is the instant at which the field is changed instantaneously. In the case described above the speed at which field changes may be made is limited by the time constant of the electromagnet and its associated circuit. When the magnetic field differs from its final steady value by less than an amount determined by the characteristics of the material, magnetic viscosity will proceed as indicated by equation (1) with the time origin at t_0, determined by the time constant of the magnet and the characteristics of the specimen. Thus equation (4) becomes

$$E(t) = -\frac{4\pi NAS_0}{t-t_0}. \qquad \ldots\ldots(5)$$

Determinations of the magnetic viscosity parameter q discussed in § 1 were made using equation (3). There are difficulties associated with the ballistic measurement of χ_{irr} involved in (3) in that for materials exhibiting magnetic viscosity the measured values depend on the time constant of the instrument used to determine the flux linking a search coil wound on the specimen (Street, Woolley and Smith 1952). These difficulties were overcome satisfactorily in the work described here by measuring the flux linkage with a Grassot flux meter having a time constant of several minutes.

§ 3. EXPERIMENTAL TESTS OF THE METHOD

The accuracy and performance of the method described above were checked by making measurements on specimens for which the magnetic viscosity parameters could also be determined by standard magnetometric techniques. Suitable specimens of Alnico 3·5 cm long and 0·6 cm in diameter were used, and for the electromagnet measurements the plane ends were machined flat and parallel. Single specimens surrounded by a coil of 6000 turns were then placed in the gap of an electromagnet; the e.m.f. induced in the coil was recorded directly on a recording potentiometer. Typical results, showing $1/E(t)$ plotted as a function of t, are given in figure 1, and are in agreement with the form predicted by (5). The specimens were then carefully ground into the form of ellipsoids, and

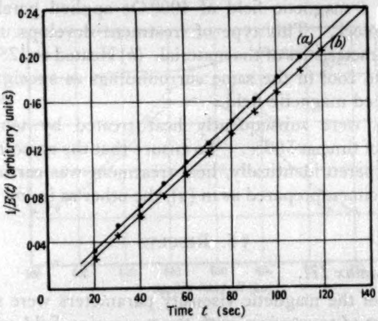


Figure 1. Typical curves showing the variation of $1/E(t)$ as a function of time for rod specimens of Alnico. (*a*) without and (*b*) with a block of Alcomax also held in the magnet gap.

magnetometer measurements of the magnetic viscosity were made by means of a fluxgate (Street, Woolley and Smith 1952). The values of S_0 and q/kT for the electromagnet method of measurement, obtained from the results given in figure 1, are 19·5 gauss and 0·59 Oe^{-1} respectively which are in good agreement with the corresponding values 19·8 gauss and 0·61 Oe^{-1} respectively determined by the magnetometer method. The values of t_0 for the small diameter rod specimens of Alnico were generally appreciably smaller than the t_0 values obtained in the measurements of Alcomax block specimens having greater cross-sectional areas. There are two reasons for this, firstly the time constant of the electromagnet and circuit varies with the cross section of the specimen in the gap and secondly differences in t_0 values arise due to the different magnetic viscosity parameters. The effect of the variation in the time constant of the magnet was simply demonstrated as follows. Two sets of measurements were made of the time dependence of $E(t)$ for an Alnico rod, firstly alone in the gap of the magnet and secondly with a block of Alcomax also included between the poles of the magnet. The two sets of results are given in figure 1, from which the increased value of the intercept t_0, due to the effect of the Alcomax, is immediately obvious. Within the limits of experimental error the slopes of the two lines are equal.

§ 4. SPECIMENS

The specimens used were of Alcomax III having a nominal weight percentage composition 24 Co, 13·5 Ni, 8 Al, 3 Cu, 0·8 Nb, remainder Fe. Alcomax III is one of a group of high coercivity alloys in which optimum permanent magnet properties may be developed by suitable heat treatment. At temperatures of 1250°c the material is in equilibrium in a single phase state but at temperatures below about 700°c the equilibrium state is at least two phase. The single phase state may be almost entirely retained by rapid cooling from 1250°c and the degree of precipitation may be progressively increased by maintaining the temperature at about 600°c for several hours.

Blocks of Alcomax III of dimensions 3·7 cm × 3·3 cm × 0·92 cm were cast and ground with accurately flat and parallel faces. The specimens were given two types of initial heat treatment: (*a*) Heated to 1250°c for 20 minutes and then allowed to cool in a magnetic field of 4000 Oe applied parallel to the longest dimension of the block. This type of treatment develops uniaxial anisotropy in the magnetic characteristics of the material. (*b*) Heated to 1250°c for 20 minutes and then allowed to cool in the same surroundings as specimens treated in (*a*) but with zero applied magnetic field.

The specimens were subsequently heat treated by tempering them for increasing periods of time at 560°c. To ensure that the anisotropic and isotropic materials were tempered identically, heat treatment was carried out on pairs of specimens, one specimen prepared as in (*a*), the other as in (*b*).

§ 5. RESULTS

(*a*) *Anisotropic Alcomax III.*

Measurements of the magnetic viscosity parameters were made on samples, after various degrees of tempering, with the measuring field applied parallel and perpendicular to the preferred axis (i.e. the direction of the field applied during cooling from 1250°c). Magnetic viscosity effects were always apparent in the measurements made with the field parallel to the preferred axis but were immeasurably small in the perpendicular direction. (It should be noted from figure 4 that measurements were made only on specimens which were reasonably well developed, having coercivities greater than 250 Oe). The sensitivity of the apparatus was such that values of S_0 in equation (5) greater than 1·5 gauss could be detected. Thus in all cases S_0 for measurements made perpendicular to the preferred axis was less than 1·5 gauss. The magnitude of irreversible magnetic viscosity effects is dependent on the number of irreversible domain processes occurring in a material. Thus the above observation is consistent with the view that for anisotropic materials, magnetization proceeds by irreversible domain processes with the field applied along the preferred axis but by reversible rotational processes when the applied field is perpendicular to the preferred direction.

Figure 2 shows a typical set of results of $1/E(t)$ plotted against t, with applied field H as a variable parameter, using an anisotropic specimen of Alcomax III. From these curves S_0 may be calculated using equation (5) and the variation of S_0 as a function of H is shown in figure 3; also shown on this figure are the associated values of χ_{irr} determined experimentally, and q, calculated from equation (3). The observed variation of q with applied field is relatively small; the variation

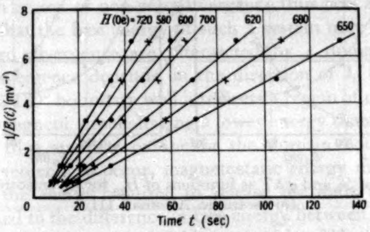


Figure 2. Typical curves showing the time variation of $1/E(t)$ with applied field H as parameter, using anisotropic Alcomax III. (Arbitrary origin of time scale.) Number of turns on search coil = 7500.

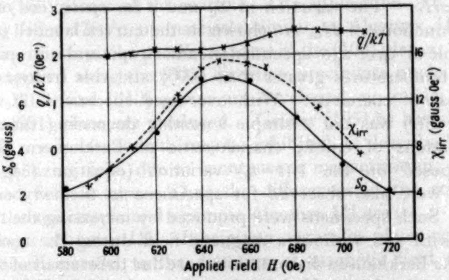


Figure 3. Variation of S_0, χ_{irr} and q/kT as functions of applied field H for anisotropic Alcomax III at temperature 293°K. Values of S_0 determined from results shown in figure 2.

of q/kT in figure 3 is partly due to the difficulty of obtaining the values of S_0 and $\chi_{\rm irr}$ under the same conditions for a given specimen. This is a difficulty which applies in all cases when the intensity of magnetization I is a function both of applied field H and time t, and has already been discussed (Street, Woolley and Smith 1952). A similar difficulty applies to the determination of the hysteresis loop and hence the coercivity H_c of a material showing magnetic viscosity. For this reason H_m, the value of H at which the maximum in the (S_0, H) curve occurs, has been used as the magnetic field parameter in presenting the results. In general it has been found that $H_m \simeq H_c$.

The curves marked (*a*) in figure 4, in which S_0 and q are plotted as functions of H_m, show the results obtained for anisotropic Alcomax III at various stages of tempering at 560°c. Similar measurements were also made on specimens of Alcomax III which had been cooled from the melt in such a way as to induce columnar growth of the crystallites.† The results obtained were very similar to those shown in figure 4, curves (*a*).

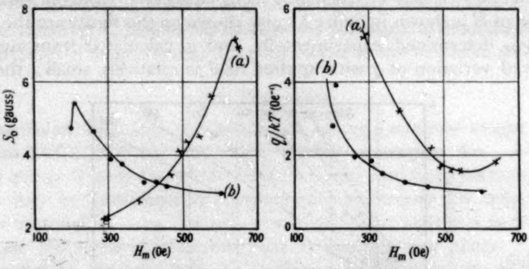


Figure 4. Variation of S_0 and q/kT as functions of H_m for (*a*) anisotropic Alcomax III and (*b*) isotropic Alcomax III.

(*b*) *Isotropic Alcomax III.*

The same sequence of measurements as described above was carried out using isotropic Alcomax III. Again for a specimen of given heat treatment, the value of q varied little with applied field and the maximum of S_0 occurred at H_m with $H_m \simeq H_c$. The variation of S_0 and q for specimens of various heat treatments as functions of H_m are shown in the curves labelled (*b*) in figure 4.

It was possible to 'over-age' specimens of anisotropic and isotropic Alcomax III by heating at temperatures greater than 600°c and this treatment resulted in a rapid decrease of coercivity. With over-aged specimens ($H_c \simeq 180$ Oe) the induced e.m.f. $E(t)$ was not a simple smoothly decreasing function of time: random fluctuations of e.m.f., characteristic of Barkhausen discontinuities, were superimposed on the $1/(t-t_0)$ variation (equation (5)). Barkhausen discontinuities were also observed for specimens in the very early stages of precipitation. Such specimens were produced by increasing the rate of cooling from 1250°c, with and without a magnetic field during the cooling. If it is assumed that the Barkhausen discontinuities are due to reversals of the magnetization vector in small volumes V, the orders of magnitude for V may be calculated from the amplitudes of the fluctuations in the e.m.f. No systematic investigations were carried out but from the results obtained it is apparent that a significant number of discontinuous changes of magnetization occur in volumes of material $V=10^{-10}$ cm^3.

---

† This type of material is known commercially as 'Columax'.

§ 6. DISCUSSION

The mechanism responsible for the production of high coercivity in Alnico-type alloys and the explanation of magnetic viscosity in terms of this mechanism have already been discussed in detail (Phillips, Street and Woolley 1954). Briefly, it is supposed that the precipitate developed by tempering is coherent with the matrix if the precipitate nuclei are not too large (Geisler 1949, 1951 a, 1953). As the precipitate has a tetragonal structure, the crystal anisotropy in the precipitate will be different from that of the matrix. Also localized regions of large stress will be produced in the vicinity of the precipitate nuclei. Although there is no general agreement as to the most important mechanism involved (Geisler 1949, 1951 a, 1953, Heidenreich and Nesbitt 1952, Nesbitt and Williams 1955), it is clear that local deviation of the spontaneous magnetization vector $\mathbf{J}_s$ away from the crystallographic axes of easy magnetization of the matrix will be produced. Thus localized regions for which div $\mathbf{J}_s \neq 0$ will be created by precipitation. The regions of non-zero divergence occur in associated pairs of positive and negative divergence so that div $\mathbf{J}_s$ over macroscopic regions of the specimen is always zero. Each region of non-zero divergence thus acts as a unipole. Néel (1946) has shown that the free energy of such a system may be decreased if the regions of non-zero divergence are intersected by a domain boundary wall. The sign of the divergence depends on the direction of $\mathbf{J}_s$ in the surrounding matrix and thus if a 180° boundary wall intersects a region of non-zero divergence it then becomes a magnetic dipole having a lower energy than the unipole. The precipitate nuclei thus stabilize or anchor the domain boundary walls. For boundary wall movements to occur, magnetostatic energy must be supplied by the external field, the energy required to release the walls from the anchored positions being equal to the difference in free energy between the dipolar and the unipolar distributions. The critical field required is taken as the coercive field.

Investigations have been made by Geisler (1951 b) and Heidenreich and Nesbitt (1952) of the physical form of the precipitate in Alnico-type materials. It has been shown that the precipitate is in the form of plates having their planes parallel to the (100) planes of the matrix. The application of a magnetic field during cooling from 1250°c of materials similar to Alcomax III inhibits the formation of precipitate nuclei parallel to some of the (100) planes. For example, with the field applied along the [001] direction precipitate nuclei having planes parallel to the (100) and (010) planes are developed, no nuclei having planes parallel to the (001) plane are produced. Geisler (1951 b) has also shown that in Alcomax III precipitate growth may be inhibited, as described above, by the internal demagnetizing field, due to spontaneous magnetization, even when the external applied field is zero. Thus the differences in the structure of the precipitate of the anisotropic and isotropic Alcomax III are:

(*a*) *Anisotropic Alcomax* III. Precipitate occurs in plates containing (100) planes most nearly parallel to the direction of magnetic field applied during cooling.

(*b*) *Isotropic Alcomax* III. Precipitate occurs in plates containing the (100) planes most nearly parallel to the local direction of the spontaneous magnetization vector during cooling. Since Alcomax III is equiaxed, i.e. there is a random distribution of the crystal axes of the crystallites, the planes of the plates of the precipitate will be oriented at random.

Alnico of nominal weight percentage composition Fe 54, Al 10, Ni 18, Co 12, Cu 6 also exhibits precipitation in plate form, but in this case precipitation is little affected by magnetic fields. Thus plates of precipitate are formed parallel to all three (100) planes in a crystallite and in equiaxed material there will be a random distribution of plates of precipitate. Accepting the view that all the magnetic characteristics of high coercivity material are related to the form and effect of the precipitate nuclei it is to be expected that the magnetic viscosity behaviour of isotropic Alcomax III should be similar to that of Alnico. The observed variations of S_0 and q, shown in figure 2 (curves (*b*)), are very similar in form to those of Alnico (cf. Phillips, Street and Woolley 1954). It has been shown that the observations of magnetic viscosity in Alnico may be qualitatively explained in terms of 180° boundary wall movements impeded by the effect of precipitate nuclei and this explanation may be directly applied to the case of isotropic Alcomax III.

The influence of anisotropy on magnetic viscosity may be analysed in terms of the activation energy theory as follows.

Consider the movement of a section of area A of a 180° boundary wall under the action of an applied magnetic field. Strictly it should be assumed that the boundary wall is deformable (Néel 1946) but for simplicity a plane boundary wall moving in a direction perpendicular to its plane will be considered here.

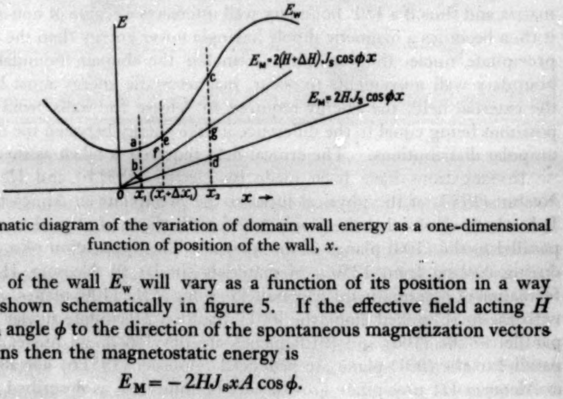


Figure 5. Schematic diagram of the variation of domain wall energy as a one-dimensional function of position of the wall, *x*.

The energy of the wall E_w will vary as a function of its position in a way similar to that shown schematically in figure 5. If the effective field acting H is inclined at an angle ϕ to the direction of the spontaneous magnetization vectors $\mathbf{J}_s$ of the domains then the magnetostatic energy is

$$E_M = -2HJ_s xA\cos\phi.$$

The condition for equilibrium of the wall is that

$$\frac{\partial}{\partial x}\{E_w - 2HJ_s xA\cos\phi\} = 0$$

and the critical field for irreversible displacement is

$$H_c = \frac{1}{2J_s A\cos\phi}\left(\frac{\partial E_w}{\partial x}\right)_{\max}.$$

$(\partial E_w/\partial x)_{\max}$ is the maximum of $(\partial E_w/\partial x)$ and occurs at x_2, the point of inflection c on the (E_w, x) curve. The equilibrium position of the wall, for effective field H, will be at x_1, defined by $(\partial E_w/\partial x)_{x_1} = 2J_s x_1 A \cos\phi H$.

In order to activate a discontinuous change in intensity of magnetization sufficient energy is required to move the wall from x_1 to x_2. The activation energy required in field H is cd − ab. If the appiied field is increased to $H + \Delta H$ the activation energy is reduced to cg − ef. Thus the change in activation energy produced by a field increment ΔH is

$$\begin{aligned}\Delta E &= (\mathrm{cg}-\mathrm{ef})-(\mathrm{cd}-\mathrm{ab}) = (\mathrm{ab}-\mathrm{ef})-\mathrm{dg}\\ &= [(E_w)_{x_1} - 2HJ_s A x_1 \cos\phi] - [(E_w)_{x_1} + (\partial E_w/\partial x)_{x_1}\Delta x_1\\ &\quad - 2(H+\Delta H)J_s(x_1+\Delta x_1)\cos\phi] - 2\Delta H A J_s x_2 \cos\phi\\ &\simeq -2\Delta H J_s(x_2-x_1)A\cos\phi - \left(\frac{\partial E_w}{\partial x}\right)_{x_1}\Delta x_1 \qquad \ldots\ldots(6)\end{aligned}$$

to the first order of small quantities. Δx_1 is the reversible movement of the boundary wall, which occurs when the field increases from H to $H + \Delta H$, and determines the reversible susceptibility χ_{rev} of the material. For the alloys considered here $\chi_{\mathrm{rev}} \ll \chi_{\mathrm{irr}}$ and thus the second term of equation (6) may be assumed zero. Hence

$$\Delta E = -2J_s v \cos\phi \Delta H$$

where $v = A(x_2 - x_1)$ = volume through which the boundary wall must move before an irreversible change in magnetization occurs.

Suppose that the specimen contains N regions of average volume $\bar{v} = \overline{A(x_2 - x_1)}$ distributed so that N_ϕ lie in boundary walls separating domains for which the spontaneous magnetization vectors are inclined at angles between ϕ and $\phi + d\phi$, with respect to the direction of measurement. Then $N_\phi = n_\phi \sin\phi d\phi$, where n_ϕ = number per unit solid angle. The total change in activation energy for the whole assembly is

$$\begin{aligned}\sum_N \Delta E &= -\sum_N 2J_s\bar{v}N_\phi \Delta H \cos\phi\\ &= -2J_s\bar{v}\Delta H \int_0^{\pi/2} n_\phi \cos\phi \sin\phi \, d\phi.\end{aligned}$$

For isotropic Alcomax III n_ϕ is a constant $= n$, independent of ϕ. The average decrease in activation energy per unit increase in applied field is, by definition, q and hence

$$\begin{aligned}q_{\mathrm{iso}} = \frac{\sum_N \Delta E}{N \Delta H} &= \frac{2J_s\bar{v}\int_0^{\pi/2}\cos\phi\sin\phi\, d\phi}{\int_0^{\pi/2}\sin\phi \, d\phi}\\ &= J_s\bar{v}.\end{aligned}$$

In the case of anisotropic Alcomax III n_ϕ will not be independent of ϕ; the spontaneous magnetization vectors will lie along the [100] directions which are nearest to the direction of the field applied during cooling. If the specimen contains N crystallites oriented crystallographically at random then the number of [100], [010] and [001] directions making angles between ϕ and $\phi + d\phi$ with respect to a fixed direction OX is $3N \sin\phi \, d\phi = -3N d\alpha$ where $\alpha = \cos\phi$ (Hoselitz and McCaig 1949). Of these the number which are *nearest* [100] etc. directions is

$$W = 3Nf(\alpha)\, d\alpha$$

where the function $f(\alpha)$ has the following values:

$$1>\alpha>1/\sqrt{2}:\ f(\alpha)=1$$

$$1/\sqrt{2}>\alpha>1/\sqrt{3}:\ f(\alpha)=1-(4/\pi)\cos^{-1}\{\alpha(1-\alpha^2)^{-1/2}\}=F(\alpha)$$

$$1/\sqrt{3}>\alpha>0:\ f(\alpha)=0.$$

Thus similarly to the isotropic case:

$$q_{\text{anis}}=\frac{2J_s\bar{v}\int_1^0 \alpha f(\alpha)d\alpha}{\int_1^0 f(\alpha)d\alpha}$$

$$=\frac{2J_s\bar{v}\left[\int_1^{1/\sqrt{2}}\alpha d\alpha+\int_{1/\sqrt{2}}^{1/\sqrt{3}}\alpha F(\alpha)d\alpha\right]}{\left[\int_1^{1/\sqrt{2}}d\alpha+\int_{1/\sqrt{2}}^{1/\sqrt{3}}F(\alpha)d\alpha\right]}.$$

The integrals involving $F(\alpha)$ may be evaluated graphically and hence $q_{\text{anis}}=2J_s\bar{v}\times 0{\cdot}831$ and the ratio $q_{\text{anis}}/q_{\text{iso}}=1{\cdot}66$ assuming that $\bar{v}$ is the same for the two materials. Qualitatively the results shown in figure 4 are in agreement with the theoretical prediction that $q_{\text{anis}}>q_{\text{iso}}$ (for constant H_{m}). It is plausible to assume that when both isotropic and anisotropic materials have been developed to have maximum H_{m} (and H_{c}) the volumes $\bar{v}$ are of optimum size, determined only by the metallurgical properties of the material. For example, the optimum size of $\bar{v}$ may be reached when the precipitate nuclei have developed to a maximum size beyond which coherence between precipitate and matrix breaks down. It is, therefore, reasonable to compare the theoretical prediction given above with the experimentally observed values of q in the region of maximum H_{m}. In this region the ratio $q_{\text{anis}}/q_{\text{iso}}$ has values between 1·5 and 2·0 which compare as well as can be expected with the predicted value of 1·66.

Theoretical interpretation of the observed variation of S_0 for isotropic and anisotropic Alcomax III is more difficult. S_0 depends on the average volume through which the boundary walls move after activation (i.e. the volume involved in a Barkhausen discontinuity) and thus S_0 is not directly dependent on $\bar{v}$ considered above. Thus detailed information on the metallurgical state of the material as precipitation proceeds is required for any attempt to explain the observed variation of S_0 as a function of H_{m}. Such information is not at present available. Qualitatively figure 4 shows obvious differences in the (S_0, H_{m}) curves for anisotropic and isotropic materials. For isotropic Alcomax III maximum S_0 occurs at low values of H_{m} and the maximum value of H_{m} is reached considerably later in progressive heat treatment. The (S_0, H_{m}) curve is similar in form to those previously observed for the isotropic materials Alnico, FeNiAl, Fe_2NiAl, PtCo etc. (Phillips, Street and Woolley 1954). For anisotropic Alcomax III and also for Columax, maximum S_0 occurs at almost the same time as maximum H_{m}. It is possible that the curves (*a*) and (*b*) of figure 4 are characteristic of anisotropic and isotropic materials respectively.

ACKNOWLEDGMENTS

The authors wish to thank Professor L. F. Bates for his continued encouragement. Their thanks are also due to Mr. Gould of the Permanent Magnet Association for the provision of the specimens used in this work, and to Dr. J. H. Phillips and Mr. D. H. Lee for their invaluable assistance during the course of the experiments. One of the authors (R. S.) acknowledges with thanks a grant from the Royal Society for the purchase of apparatus used in this work.

REFERENCES

GEISLER, A. H., 1949, *Trans. Amer. Inst. Min.* (*Metall.*) *Engrs*, **180**, 230; 1951 a, *Phase Transformations in Solids*, Ed. R. Smoluchowski (New York: Wiley), p. 387; 1951 b, *Phys. Rev.*, **81**, 478; 1953, *Rev. Mod. Phys.*, **25**, 316.

HEIDENREICH, R. D., and NESBITT, E. A., 1952, *J. Appl. Phys.*, **23**, 352.

HOSELITZ, K., and MCCAIG, M., 1949, *Proc. Phys. Soc.* B, **62**, 163.

MEE, C. D., and STREET, R., 1954, *Proc. Instn Elect. Engrs*, Pt. II, **101**, 639.

NÉEL, L., 1946, *Ann. Univ. Grenoble*, **22**, 299; 1950, *J. Phys. Radium*, **11**, 49; 1951, *Ibid.*, **12**, 339.

NESBITT, E. A., and WILLIAMS, H. J., 1955, *J. Appl. Phys.*, **26**, 1217.

PHILLIPS, J. H., STREET, R., and WOOLLEY, J. C., 1954, *Phil. Mag.*, **45**, 505.

SIMPSON, A. W., and TREDGOLD, R. H., 1953, *Proc. Phys. Soc.* B, **66**, 805.

SNOEK, J. L., 1947, *New Developments in Ferromagnetic Materials* (Amsterdam: Elsevier).

STREET, R., and WOOLLEY, J. C., 1949, *Proc. Phys. Soc.* A, **62**, 562.

STREET, R., WOOLLEY, J. C., and SMITH, P. B., 1952, *Proc. Phys. Soc.* B, **65**, 679.

Changes in Domain Structure accompanying Reduction in Depth of a Single Crystal of Silicon Iron

By L. F. BATES and A. HART†

Physics Department, University of Nottingham

MS. received 13*th June* 1956

Abstract. Bitter patterns reveal marked changes in the domain structure of a Néel-cut crystal of silicon–iron when the crystal dimension perpendicular to a (110) plane is sufficiently reduced, and suggest the formation of a new structure with an energy comparable with that of the normal Néel type. Such changes in the domain configuration must be accompanied by corresponding changes in magnetic properties.

---

§ 1. Introduction

In an earlier publication (Bates and Hart 1956) Bitter patterns observed on the (01Ī) face of a Néel-cut crystal (figure 1) magnetized in a [011] direction were reported. These powder pattern experiments combined with others described by Bates and Neale (1950), Bates and Mee (1952), showed that for a single crystal with a (01Ī) plane roughly of dimensions 1·0 cm by 0·6 cm, the depth L of the crystal in the [01Ī] direction perpendicular to the (01Ī) plane did not affect the type of domain structure in the specimen, at least in the range of depths 0·6 cm to 0·065 cm. The structure deduced from the patterns approximated to that first proposed by Néel (1944) (figure 1), but was characterized on the (01Ī) surfaces by the familiar 'lace' patterns (Bates and Mee 1952). The present paper deals with patterns observed on the same specimen in the range of depths between 0·065 cm and 0·025 cm, which are markedly different from those found on the thicker specimens.

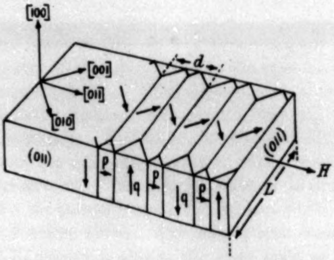


Figure 1. Domain structure in a Néel block.

---

† Now at Post Office Research Station, Dollis Hill, London, N.W.2.

§ 2. EXPERIMENTAL ARRANGEMENT AND RESULTS

The patterns were observed on the $(01\bar{1})$ plane of a specimen of 3·8% silicon–iron of dimensions 1·25 cm by 0·58 cm. The depth of the specimen L perpendicular to the $(01\bar{1})$ plane was initially 0·065 cm, but during the course of the experiments it was reduced by successive mechanical and electrolytic polishings to 0·025 cm. During the observations the ends of the specimen were held in light contact with the pole faces of an electromagnet. The surface under observation was always electrolytically polished, and the patterns were viewed through a metallurgical microscope with bright field illumination. Before observation of the patterns the specimen was first put into a cyclic state by reversing the magnetizing field some twenty times.

Figure 2 (Plate) shows some 'normal' lace patterns recorded when the specimen depth was 0·065 cm. The magnetizing field was applied parallel to the long axis of the specimen—i.e. parallel to the [011] direction. The alternate light and dark bands of the pattern run at right angles to this long axis, and are thus parallel to a [100] direction; the effective field was approximately 10 Oe. When the specimen depth was reduced to 0·06 cm the bands of the pattern remain parallel to each other but were now inclined at a small angle (about 5°) to the [100] direction, the effective field being unchanged.

Increasing the field caused the bands to become parallel to the [100] direction, but reduction of the field after reaching a high value was always accompanied by a deviation of the bands from that direction. The observed angular deviations became larger as the specimen depth was further reduced, until at a depth of 0·035 cm the bands of the pattern in a low field ran at an angle of nearly 30° to their usual [100] direction, as in figure 3 (Plate). It is emphasized at this point that the specimen on which these patterns were recorded had its main surfaces ground to within 0·25° of $(01\bar{1})$ planes, and that its long axis was less than 1° from a true [011] direction.

At a depth of 0·035 cm, and in an effective magnetic field of about 10 Oe, the characteristic bands of the lace pattern were present on isolated sections of the surface only. Over the rest of the specimen they were replaced by a series of lines approximately parallel to each other and parallel to the long axis of the specimen, i.e. these lines were perpendicular to the bands of the original lace pattern. A photograph of these new patterns is given in figure 4 (Plate), and comparison with figure 2 shows what marked changes have occurred. Incidentally, figure 4 was obtained with a small field applied normal to the specimen surface, in order to accentuate alternate lines of the pattern and to make them more easily visible.

The lines of the new pattern appear to be less rigidly restricted to a particular crystal direction than do most simple line patterns which have previously been reported. This characteristic is most obvious in the transition regions between the new patterns and the isolated islands of lace pattern, figure 6 (Plate). There appears to be complete continuity between the two types of pattern, and to maintain this continuity some pattern lines must turn through angles of up to 70°.

As the applied field was increased the lines broadened and assumed a chain-like appearance as in figure 5. Further increase in field caused the chains to become wider, and finally in fields greater than 200 Oe there was a reversion to normal lace patterns running parallel to the [100] direction.

In the presence of a small field normal to the plane of the specimen, as already mentioned, alternate lines of the new pattern were accentuated, figure 4, and reversal of the normal field caused the usual interchange of light and heavy deposits. A larger normal field caused alternate lines of the original pattern to appear as 'chains'. The line and chain deposits exchanged positions on reversal of the normal field. At still smaller specimen depths the regions over which normal lace patterns could be seen were even more restricted, and at 0·025 cm the whole specimen was characterized by the new chain pattern.

§ 3. Proposed Domain Structure

The changes in the patterns as the specimen depth was reduced indicate that the transition, between lace patterns and the new ones, is a gradual one closely related to the depth of the specimen. A structure which accounts for the patterns observed in the thinnest specimens, representing the culmination of the transition, will now be studied. This structure holds strictly for zero field and is shown in figure 7.

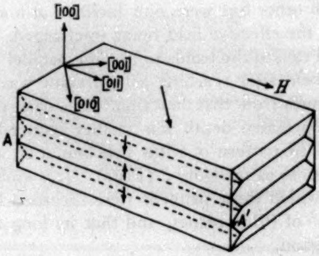


Figure 7. Domain structure in a thin specimen.

The bulk of the specimen is still occupied by domains magnetized alternately along the [010] and [001] directions but the walls between domains are in the (100) rather than the (011) planes. This is possible since the requirement that the normal component of magnetization across the boundary must be continuous is satisfied by any one of the zone of planes having the [01$\bar{1}$] as zone axis. Flux closure at the (01$\bar{1}$) surface is achieved by domains of triangular cross section running the whole length of the specimen and magnetized alternately in the [100] and [$\bar{1}$00] directions. Using this model the lines of the low-field pattern represent intersections on the (01$\bar{1}$) plane of boundaries such as AA′ in figure 7.

As the applied field is increased the component of the main domain magnetization perpendicular to the (01$\bar{1}$) plane decreases, and the need for complete flux closure via the triangular closure domains may be relaxed sufficiently to allow a shrinking of the closure domains, accompanied by the appearance of strips of free poles between closure domains. It is suggested that these strips give rise to the 'chain' pattern, and that this pattern is thus very similar in origin to a normal lace pattern.

The effects of both large and small normal fields on the patterns can be explained by reference to figure 7. A large normal field causes a rotation of the main domain magnetization vectors, neighbouring domains suffering respectively an increase and a decrease in their components of magnetization perpendicular to the $(01\bar{1})$ plane. By virtue of their reduced normal component alternate domains will then be able to approach the $(01\bar{1})$ surface via a free pole (chain) structure, while their neighbours will still retain the closure domain system. Alternate lines and chains are, in fact, observed under these conditions. Also, when a strong normal field is applied to a specimen exhibiting the chain pattern, alternate chains are suppressed. As might be expected from the foregoing argument, increasing the magnetizing field parallel to the [011] causes a reduction in the component of magnetization normal to the $(01\bar{1})$, and therefore allows the suppressed chains to reappear.

§ 4. ENERGY CONSIDERATIONS

The excess energy of the proposed structure, comprising the anisotropy energy of the closure domains and the wall energy of the main domains, may be minimized to obtain an estimate of the equilibrium domain spacing. The analysis is similar to that of Néel for his original structure and will not be reproduced here. It is emphasized that the theoretical calculations approximate most closely to the experimental cases in low fields, and that in higher fields, where the patterns indicate that a considerable fraction of the main domains approach the surface by way of a structure involving free poles, the calculations are limited in the same manner as are those of Néel when the p domains of figure 1 are replaced by lace patterns.

The calculations show that at all depths the equilibrium domain spacing of the new structure is roughly 30% greater, at 5 Oe, than the corresponding spacing in the Néel structure, and that it remains greater up to fields near 200 Oe. The volume of the new type of closure domains at all field strengths is less than the volume of the corresponding q domains of the Néel structure, figure 1, while the excess energy of the new domain structure is less than that of the old structure in all fields up to 300 Oe. Thus the so-called ‘normal’ lace pattern should only occur at fields greater than 300 Oe while at all fields less than this the new pattern should be observed. This result is supported qualitatively by the patterns recorded on the thinner specimens, but not by those on the thicker ones. This leads us to the conclusion that in the above-mentioned calculations an energy term has been ignored which is sufficiently important to influence the energy balance between the two structures, making the normal lace structure more favourable at all fields for the thicker specimens. The ratio of the excess energy of the domain structure (wall energy of main domains plus anisotropy energy of closure domains) to the anisotropy energy in the bulk of the specimen increases with $(\text{depth})^{-1/2}$ as the depth decreases. At a depth of 0·04 cm, in an effective field of 2 Oe, the difference between the excess energies of the two types of domain structure represents 8% of the bulk anisotropy energy, but at greater depths it will become considerably smaller. It seems that only in thin specimens is the gain in energy associated with the adoption of the new structure large enough to outweigh the energy term(s) which cause the lace structure to be favoured in thicker specimens at fields less than 300 Oe.

In any event, the gradual change in the patterns and the fact that structures such as those indicated by the patterns of figure 6 can exist, indicate that the energies of the two structures are so nearly equal as to require very refined calculation to decide which is the more favourable. When calculations involving the largest energy terms give such similar results, smaller energy terms—for example that associated with the magnetostrictive strain energy of the domain configuration—which are difficult to assess, may be expected to play a big part.

§ 5. DISCUSSION AND CONCLUSIONS

The calculations mentioned in the previous section indicate that the domain structure proposed for thin specimens can be energetically preferable to the normal Néel structure. The obvious indications that all factors effecting the transition from one structure to the other have not been included in the calculations make it unprofitable to use energy considerations to explain the dependence of the exact form of the structure on the specimen depth. All that can be said is that there appears to be complete flux closure between the two structures, and that the transition from one to the other is gradual and closely dependent on specimen depth.

Lace patterns deviating from the [100] direction were previously reported by Bates (1954) for iron single crystals and polycrystalline specimens of grain-oriented silicon–iron. Dijkstra and Martius (1953) reported patterns very similar to those of figure 4 on grain-oriented silicon–iron polycrystalline material, subjected to stress. The lines in their patterns showed the same tendency to change direction, and the structure presented here is similar to the one proposed by them.

The change in domain structure reported in this paper is dependent on the increasing ratio of surface to volume energy attendant on the decrease in size of the specimen used. That such a change in the balance of energy can cause large changes in the observed magnetic characteristics is obvious in the two extreme cases represented respectively by single-domain particles and films. While the changes in magnetic properties accompanying the above change in domain structure may not be as pronounced as in the multi-domain to single-domain transition, they may still be important. The change in susceptibility of a Néel-cut crystal as the scale of its domain structure alters has already been calculated (Lee 1953, Bates and Hart 1956) and shown to be significant. Equally, changes in the type of the domain structure existing in a specimen will affect properties such as the susceptibility, coercivity, etc.

ACKNOWLEDGMENTS

The crystal used in these experiments was that presented by Dr. D. Shoenberg and used by Bates and Mee; one of us (A. H.) is indebted to the Department of Scientific and Industrial Research for a maintenance allowance during the whole of the work.

REFERENCES

BATES, L. F., 1954, *J. Inst. Met.*, **82**, 417.
BATES, L. F., and HART, A., 1956, *Proc. Phys. Soc.* B, **69**, 497.
BATES, L. F., and MEE, C. D., 1952, *Proc. Phys. Soc.* A, **65**, 129.
BATES, L. F., and NEALE, F. E., 1950, *Proc. Phys. Soc.* A, **63**, 374.
DIJKSTRA, L. J., and MARTIUS, U. M., 1953, *Rev. Mod. Phys.*, **25**, 146.
LEE, E. W., 1953, *Proc. Phys. Soc.* A, **66**, 623.
NÉEL, L., 1944, *J. Phys. Radium*, **5**, 241, 265.

Paramagnetic Resonance in some Complex Cyanides of the Iron Group

I : Experimental Results

By J. M. Baker, B. Bleaney and K. D. Bowers
Clarendon Laboratory, Oxford

MS. received 31*st August* 1955 *and in revised form* 19*th June* 1956

Abstract. The results are described of paramagnetic resonance experiments on some complex cyanides of the iron group. The form of the spin Hamiltonian is found and the values of the parameters are measured. Potassium chromicyanide and potassium vanadocyanide (configuration $3d^3$), diluted with the isomorphous diamagnetic salts, have effective spin 3/2 with an electronic splitting of rhombic symmetry. The *g*-value is closely isotropic at 1·992, and the hyperfine structure is also isotropic. Potassium ferricyanide and potassium manganicyanide (configuration $3d^5$), diluted with the isomorphous diamagnetic salt, have effective spin ½ with anisotropic *g*-values between 0·6 and 2·6. The manganicyanide also exhibits an anisotropic hyperfine structure. The behaviour of undiluted potassium chromicyanide and potassium ferricyanide suggests that there is appreciable exchange interaction between neighbouring paramagnetic complexes.

§ 1. Introduction

In a first approximation the paramagnetism of salts of the iron (3d) transition group is due only to the electronic spin, the orbital moment being 'quenched' by the crystalline electric field. The effective value of the spin *S* may be deduced from susceptibility measurements, and for most hydrated salts the values agree with those predicted on the basis of Russell–Saunders coupling for a configuration d^n, for which $S=\frac{1}{2}n$ or $\frac{1}{2}(10-n)$, whichever is the smaller. In particular, a value $S=5/2$ is obtained for the d^5 configuration, a half-closed shell for which $L=0$ in the free ion. Effects of the crystalline electric field are then very small, and the susceptibility of most salts of the $3d^5$ ions, Mn^{2+} and Fe^{3+}, is notable for the very small anisotropy and the accuracy with which Curie's law is followed (apart from exchange effects in magnetically concentrated salts). The compound potassium ferricyanide $K_3Fe(CN)_6$ has long been recognized as an exception; the anisotropy of the susceptibility of a single crystal was observed long ago by Grailich and von Lang (1858) who showed that the susceptibility was greatest along the *a*-axis and least along the *c*-axis. Also the susceptibility departs markedly from Curie's law (Jackson 1933, 1938, Masson 1947, Guha 1951) and its magnitude is nearer to that expected for a spin $S=1/2$ than that for $S=5/2$.

An explanation of this behaviour was first put forward by Pauling (1931) on the basis of covalent binding of the Fe^{3+} ion to the surrounding cyanide groups, and Van Vleck (1935) has shown that the same result is obtained by the use of molecular bond orbitals or by exceptionally large crystalline electric fields. For an octahedral complex the result is, effectively, that only three 3d orbital levels are available to accommodate the magnetic electrons, instead of five. The electrons go into these three levels in accordance with a modified form of Hund's rule, that the spin multiplicity has the maximum value consistent with the Pauli exclusion principle when only three orbitals are available. The effective value of S is then $\frac{1}{2}n$ or $\frac{1}{2}(6-n)$, whichever is the smaller, just as for an atomic p-shell. The experimental evidence from susceptibility measurements is in agreement with this for the complex cyanides of the iron group (Kotani 1949); in particular, compounds with the d^6 configuration are diamagnetic (or have a small temperature independent paramagnetism) and compounds of d^n with n greater than 6 are generally unstable.

In an early survey of paramagnetic resonance, Bagguley, Bleaney, Griffiths, Penrose and Plumpton (1948) showed that the spectrum of a powder of $K_3Fe(CN)_6$ was very different from that of a ferric compound with 'ionic' binding. In view of the detailed information which resonance measurements were expected to yield, it was decided to make a general survey of the complex cyanides of the iron group using single crystals. The results of this work, some of which have been briefly reported earlier, are presented in this paper. The theoretical aspects, based on the theory of the covalent XY_6 complex of Stevens (1953), are considered in the following paper (Bleaney and O'Brien 1956).

Since the immediate surroundings of the paramagnetic ion are carbon and nitrogen atoms with small or zero nuclear moments, it should be possible to obtain narrower lines on dilution than in hydrated salts (see Bowers 1952). This reduced line width was realized in compounds of $3d^3$ ions and was utilized in the determination of the nuclear spins of ^{53}Cr and ^{50}V; the results have already been published (Bowers 1952, Baker and Bleaney 1952) and will only be summarized here. In the $3d^5$ cyanides the minimum line width attained on dilution was considerably greater; since these ions have very anisotropic g-values, this is probably due to variation of the crystalline electric field in different parts of the crystal. In the concentrated salts the line widths indicate the presence of considerable exchange forces, as will be discussed in §5.

The complex cyanides which have been studied are those of the ions $3d^3$ $(S=3/2)$ and $3d^5$ $(S=1/2)$. The $3d^1$ ion (e.g. Ti^{3+}) forms no stable salt, while the $3d^2$ and $3d^4$ ions do not have Kramers degeneracy $(S=1)$ and the levels are so split up that one would not anticipate any observable transitions at centimetre wavelengths. It was verified that no spectrum could be observed from the $3d^4$ ion Mn^{3+} in $K_3Mn(CN)_6$. Both the divalent and the trivalent ions of d^3 and d^5 have been studied, the salts used being $K_4V(CN)_6 . 3H_2O$ and $K_3Cr(CN)_6$ for d^3, and $K_4Mn(CN)_6 . 3H_2O$ and $K_3Fe(CN)_6$ for d^5. The isomorphous salts of the configuration $3d^6$, which are diamagnetic, have been used as diluents; these salts are $K_4Fe(CN)_6 . 3H_2O$ and $K_3Co(CN)_6$. Where diluted salts are mentioned below, these diluents should be assumed, except in one case where an alternative is specified. The apparatus used for all measurements was the standard type described by Bleaney and Stevens (1953) with radiation at 0·3 and 0·8 cm^{-1}. Experiments were carried out at constant frequency and variable magnetic field, using single crystals at various temperatures. In the d^3 cyanides resonance can be observed at room temperature, but in the d^5 cyanides the spin–lattice relaxation time is very short and liquid hydrogen temperatures are required

§ 2. CRYSTALLOGRAPHY

The salts under review fall into two isomorphous series (see Bowers and Owen 1955), whose magnetic behaviour is rather similar. In each case the unit cell contains only two distinct magnetic complexes, each comprising a metal ion M surrounded by a nearly regular octahedron of six CN groups; the two complexes differ only in the different orientation of their magnetic axes.

The salts of trivalent ions, general formula $K_3M'''(CN)_6$, are grown from aqueous solution to which (except for Fe and Co) KCN is added. With this precaution no difficulty is found in growing crystals of the required size, up to $\frac{1}{2}$ cm cube. The morphology of the crystals is pseudo-orthorhombic (see Groth 1906), and for convenience we shall give the direction cosines of the relevant axes relative to the ortho-rhombic axes *a*, *b*, *c*, though the unit cell has been shown to be monoclinic (Barkhatov 1942, Barkhatov and Zhdanov 1942). This work shows that each of the two cyanide octahedra has a two-fold axis along the *c* axis; the four-fold axis normal to the *c* axis makes an angle of 27° with the *b* axis, and the two inequivalent octahedra transform into each other by reflection in the *ac* plane. The direction cosines of the axes (α, β, γ) of the two octahedra are

	α	β	γ
a	0·45	−0·63	0·63
b	±0·89	±0·32	∓0·32
c	0	0·71	0·71

The divalent metal ions form salts of the formula $K_4M''(CN)_6 . 3H_2O$ and are grown from aqueous solution where, even with considerable excess KCN, all except the diamagnetic ferrocyanide are rather unstable. No large crystals except of the latter could be grown, but, by using air-free water and a vacuum desiccator technique by which the crystals could be grown in a few hours, mixed crystals of the ferrocyanide with about 1% of the required paramagnetic salt were obtained as plates about 4 mm square and 1 mm thick. The morphology and crystal structure are given respectively by Groth (1906) and Pospelov and Zhdanov (1947). There are two types of crystal; one is monoclinic, pseudo-tetragonal, with lattice constants $a=c=9{\cdot}34$Å, $b=16{\cdot}86$Å, $\beta=90°\pm5'$, four molecules in unit cell and space group C_{2h}^{6}; the other is tetragonal with constants $a=b=9{\cdot}37$Å, $c=33{\cdot}69$Å, and space group C_{4h}^{6}. This has eight molecules in its unit cell, which essentially comprises two unit cells of the monoclinic type contiguous in their *ac* planes but with one rotated through 90° from the other; thus the unit cell is twice as large in one dimension. The monoclinic type shows a tendency to form 90° twins in the *ac* plane. The crystals grow in flat rectangular plates parallel to the *ac* plane, thus having well developed 010 faces. The other principal faces are 110 and 011 which form the edges of the plates, so that there is no way of distinguishing between the *a* and *c* axes by visual inspection.

The x-ray data on the monoclinic type show that all the cyanide octahedra have the same orientation; one axis of the octahedron points along the *b* axis, and the other two make angles of 18° with the *a* and *c* axes. The direction cosines of the octahedral axes α, β, γ with respect to the monoclinic axes (taking the monoclinic angle as 90°) are as follows

	α	β	γ
a	0	0·95	−0·31
b	1	0	0
c	0	0·31	0·95

Magnetically two complexes are found, as expected with C_{2h} symmetry, one derived from the other by reflecton in the *ac* plane; this shows that there must be distortions of the cyanide octahedra not revealed by the x-ray analysis. Also, crystals which are obviously twinned exhibit four distinct magnetic complexes instead of two. Twins of the tetragonal type, involving 90° rotation about the *b* axis, would show the same number of magnetic complexes as untwinned crystals, since the tetragonal variety has 90° rotational symmetry about this axis. Hence the untwinned crystals, on which the resonance data reported below were obtained, must be of the monoclinic variety.

§ 3. EXPERIMENTAL RESULTS ON $3d^3$ COMPLEXES

It is found that the paramagnetic resonance spectra of the complexes $K_3Cr(CN)_6$ and $K_4V(CN)_6 . 3H_2O$ can be interpreted in terms of the spin Hamiltonian:

$$\mathscr{H} = \beta \mathbf{H} . g . \mathbf{S} + D\{S_z^2 - \tfrac{1}{3} S(S+1)\} + E(S_x^2 - S_y^2) + A\mathbf{S}.\mathbf{I}. \quad \ldots\ldots(1)$$

Here β is the Bohr magneton, H the applied magnetic field, g the spectroscopic splitting factor; S (the effective spin) is 3/2, as one would expect for three unpaired electrons; D represents the effect of an initial splitting due to fields of axial symmetry, E represents the departure from this axial symmetry, and the last term represents the interaction between the electrons and the magnetic moment of the nucleus which has nuclear spin I. The directions of the magnetic axes (x, y, z) were determined experimentally and were chosen in such a way as to make the sign of E the same as that of D.

The term in A is much smaller than the others, and first these latter may be considered by themselves. When H is applied along the z axis, they give rise to four energy levels between which transitions occur at:

$$\left.\begin{array}{ll} M = 3/2 \longleftrightarrow 1/2 & G = h\nu - 2D + 3E^2D/(h\nu)^2 \\ M = 1/2 \longleftrightarrow -1/2 & G = h\nu \qquad\quad - 3E^2/h\nu \\ M = -1/2 \longleftrightarrow -3/2 & G = h\nu + 2D - 3E^2D/(h\nu)^2 \end{array}\right\} \quad \ldots\ldots(2)$$

where G is $g_z\beta H$, and $h\nu$ is a quantum of the microwave radiation. The corresponding formulae (all correct to the third order) for the x and y directions may be obtained by replacing D by $-\frac{1}{2}(D \mp 3E)$ and E by $-\frac{1}{2}(E \pm D)$, (upper signs for the x direction).

The effect of the term in A is to split each of these three transitions into $2I+1$ components. In the first approximation these are equally spaced with separation A, but there are small second order effects which depend upon the relative sign of D and A and so permit the relative signs to be found (Bleaney 1951). The overall width $\Delta\nu$ of the hyperfine structure of the transition $(M)(\longleftrightarrow M-1)$ is

$$I\,|\,[2A + A^2(2M-1)/h\nu]\,|.$$

Hence $\Delta\nu$ is greatest or least for the $3/2 \longleftrightarrow 1/2$ transition depending on whether A is positive or negative respectively. Also equations (2) show that, when H is applied along the z axis, the $3/2 \longleftrightarrow 1/2$ transition occurs in the lowest or highest field depending upon whether D is positive or negative respectively. Hence the relative signs of A and D may be determined by finding whether $\Delta\nu$ is greater for the low field line or for the high field one. The actual sign of D may be found in the following way. As the temperature falls the upper levels (which in strong magnetic fields correspond to positive values of M) have a declining population, and the intensity of the $3/2 \longleftrightarrow 1/2$ transition is reduced relative to that of the $-1/2 \longleftrightarrow -3/2$ transition.

Hence the transitions giving rise to the outer lines in the spectrum may be identified by observing their relative intensity as a function of temperature, and the sign of D is then determined. It may be remarked that even at high temperatures the two transitions may not have the same intensity: when H is applied along the z axis the fractional difference in the intensities may have any value up to approximately $4|E|/G$, the actual value depending upon the orientation of the radio-frequency magnetic field.

3.1. *Potassium Chromicyanide, Dilute*

The spectrum of potassium chromicyanide was examined at 20°K using radiation of wave number about 0·77 cm^{-1} in dilute crystals containing between 0·1 and 1% chromicyanide. The spectrum observed is due to two magnetically similar but differently oriented complexes in unit cell. The separation between the two outer lines of each complex has an extreme value (from which D, E are determined) along each of its three principal magnetic axes x, y, z, and these axes were located by examining the angular variation of the spectrum. Their direction cosines with respect to the crystallographic axes a, b, c are given in table 1. The value of g was found by comparing the position of the central line with that of the resonance line of diphenyl trinitrophenyl hydrazil ($g = 2{\cdot}0036$), and applying the correction for the second order shift indicated by equation (2). The results are given in table 1, and show that the three principal g-values are nearly equal. The stated error arises in the determination of the position of the central line, and in the correction for the second order shift (this correction is quite large, amounting to 0·021 in the x direction).

Table 1. Summary of Results for 3d^3 Complexes

Paramagnetic salt		$K_3Cr(CN)_6$			$K_3Cr(CN)_6$			$K_4V(CN)_6 . 3H_2O$		
Diluent		$K_3Co(CN)_6$			$K_3Mn(CN)_6$			$K_4Fe(CN)_6 . 3H_2O$		
D (cm^{-1})		+0·0831 ± 0·0010			+0·0538 ± 0·0010			−0·0264 ± 0·0004		
E (cm^{-1})		+0·0108 ± 0·0010			+0·0120 ± 0·0010			−0·0072 ± 0·0004		
Principal axes		x	y	z	x	y	z	x	y	z
Direction	a	0·104	0	0·994	0	0	1	0·707	0·523	0·470
cosines	b	±0·994	0	∓0·104	±0·996	∓0·087	0	0	±0·676	∓0·737
	c	0	1	0	0·087	0·996	0	0·707	−0·523	−0·470
g-value		1·993	1·991$_4$	1·991	1·992	1·995	1·993	1·9919	1·9920	1·9920
Probable error in g		±0·001			±0·002			±0·0006		
A (10^{-4} cm^{-1})		(^{53}Cr) (+) 14·7 ± 0·5						(^{51}V) −55·5 ± 0·3		
								(^{50}V) −21·1 ± 0·3		
Nuclear spin		(^{53}Cr) $I = 3/2$						(^{51}V) $I = 7/2$		
								(^{50}V) $I = 6$.		

Note: The sign of A for ^{53}Cr was not directly determined: it is inferred from the fact that the nuclear moment is known to be negative, whereas that of ^{51}V is positive.

In surroundings of rhombic symmetry the energy levels of $K_3Mn(CN)_6$ (where the Mn''' is in a d^4 state) would all be split into singlets, and no resonance spectrum would be expected at centimetre wavelengths; none was found over the temperature range 12 to 290°K. The susceptibility measurements of Cooke and Duffus (1955) show that the ground state must be a singlet; in a field of

H kilogauss this acquires only a small induced moment of order $(0{\cdot}004\,H)\beta$. Thus this salt should act as a suitable diluent for the chromicyanide, since at low temperatures the induced moments on the Mn ions would only give a small shift in the resonance field of the Cr ion, while at higher temperatures ($\simeq 100°\text{K}$), where the susceptibility measurements indicate that other levels are being populated in the manganese complex, a small broadening of the Cr lines would be expected. These expectations were confirmed, and the results using manganicyanide as diluent are also given in table 1. The principal difference from the previous results using cobalticyanide as diluent is the smaller value of the splitting constant D.

3.2. *Potassium Vanadocyanide, Dilute*

The spectrum of potassium vanadocyanide was examined at 90°K using radiation of about $0{\cdot}8\,\text{cm}^{-1}$ in crystals which probably contained between 0·1 and 5% vanadocyanide. The exact quantity could not be controlled because of the rapid oxidation of the salt in solution. The untwinned crystals exhibited two magnetically similar but differently oriented complexes in unit cell; the principal axes are given in table 1, together with the values of the parameters g, D and E, determined in the same way as for the chromicyanide. As the value of D is much smaller than in the chromicyanide the correction to the g-value due to the second order shift is much smaller, and the g-value is seen to be isotropic within experimental error. The sign of D was found to be negative.

§ 4. EXPERIMENTAL RESULTS ON 3d^5 COMPLEXES

It is found that the paramagnetic resonance spectra of each of the complexes, $K_3Fe(CN)_6$ and $K_4Mn(CN)_6 . 3H_2O$, can be interpreted in terms of a spin Hamiltonian :

$$\mathcal{H} = \beta \mathbf{H} . g . \mathbf{S} + \mathbf{S} . A . \mathbf{I}. \qquad \ldots\ldots(3)$$

Here the effective spin $S = 1/2$, as would be expected for a d^5 octahedral complex with covalent binding. Both the spectroscopic splitting factor g and the hyperfine constant A are anisotropic and must be considered as tensors (with the same axes). The spectrum consists of a single electronic transition from each paramagnetic complex which, in the case of the manganese salt (100% ^{55}Mn), is split into six components corresponding to the known nuclear spin ($I = 5/2$).

4.1. *Potassium Ferricyanide, Dilute*

The spectrum of diluted potassium ferricyanide was examined at 20°K using radiation of 0·3 and $0{\cdot}8\,\text{cm}^{-1}$ in crystals containing 0·1 to 1% of ferricyanide in potassium cobalticyanide. As in the chromicyanide there are two magnetically similar but differently oriented complexes in unit cell. The directions of their principal axes and the values of g along these axes are given in table 2 ; the z axis of each ion coincides with the crystallographic c axis, and the x, y axes lie in the ab plane. The spectra of the two ions are identical along the a and b axes and, of course, along the c axis.

The only stable odd isotope of iron is ^{57}Fe, natural abundance 2·2%, and no hyperfine structure could be resolved. This is not surprising, because the smallest line half width obtained was about 30 gauss, which is larger than the splitting anticipated for any hyperfine structure (compare Bleaney and Trenam 1954).

4.2. *Potassium Manganocyanide, Dilute*

The spectrum of potassium manganocyanide was examined at 12°K using radiation of both 0·3 cm^{-1} and 0·8 cm^{-1} in crystals probably containing between 0·1 and 5% manganicyanide (like the vanadocyanide, this oxidizes rapidly in solution). The untwinned crystals exhibit two magnetically similar but differently oriented complexes in unit cell. The principal axes, found by examination of the spectrum in a number of planes, are given in table 2, together with the g-values. The values of g_x and g_y were measured directly, but, as the available fields were not high enough, the value of g_z was measured by extrapolation.

The hyperfine structure is only resolved when the temperature is reduced to about 12°K. It comprises six nearly equally spaced lines. The values of the hyperfine structure constants A_x and A_y were measured directly and A_z by extrapolation in a plane containing the z axis. The values are given in table 2. As far as could be detected the principal axes of the A tensor were the same as those of the g tensor.

Table 2. Summary of Results for 3d^5 Complexes

Paramagnetic salt		$K_3Fe(CN)_6$			$K_4Mn(CN)_6 . 3H_2O$		
Diluent		$K_3Co(CN)_6$			$K_4Fe(CN)_6 . 3H_2O$		
Principal axes		x	y	z	x	y	z
Direction cosines	a	0·866	0·500	0	0·864	0·410	0·295
	b	±0·500	∓0·866	0	∓0·105	∓0·423	±0·899
	c	0	0	1	0·495	−0·807	−0·322
g		2·35	2·10	$0{\cdot}91_5$	2·624	2·182	0·63
		±0·02	±0·02	±0·01	±0·008	±0·008	±0·10
$\lvert A \rvert$ (10^{-4} cm^{-1})					84·5	46·5	104
					±0·5	±0·5	±20

At 12°K the lines reached a minimum half width in the x direction of 13 gauss at half intensity. As the crystal is rotated the lines move to higher fields and become broader; the actual breadth, however, was difficult to estimate because of the hyperfine structure and the decreasing signal strength. A similar behaviour is found in the diluted ferricyanide.

In no direction was there any sign of small lines between the six main hyperfine lines which might be attributed to interaction with a nuclear electric quadrupole moment, but as the lines were so broad there was little chance of seeing them even if they were present. In the most favourable direction, near the x axis, they would have been resolved only if their intensity had been greater than one hundredth of the main lines.

§ 5. MEASUREMENTS ON UNDILUTED SALTS

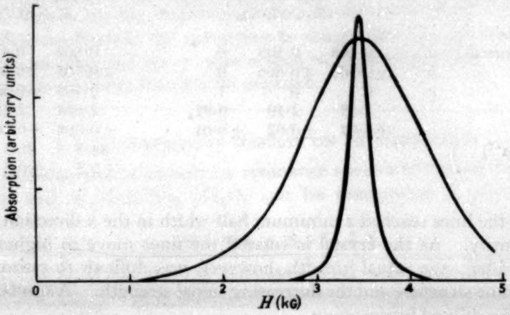
Large crystals of the undiluted salts could only be grown for the more stable compounds $K_3Cr(CN)_6$ and $K_3Fe(CN)_6$, and the results will be summarized separately.

5.1. *Potassium Chromicyanide*

The spectrum of a single crystal of $K_3Cr(CN)_6$ consisted of just one broad line whose centre was nearly isotropic in position. In one direction the position of the centre was measured as accurately as possible at two wavelengths and, assuming that any displacement is inversely proportional to ν as given by equations (2) for the dilute salts, the true g-value was estimated to be 1·998 ± 0·008.

The width and shape of the line were very anisotropic. The root mean square width h was computed from the lines observed in a number of directions at $0{\cdot}3\ \mathrm{cm}^{-1}$, and the values are given in table 3. The variation is roughly the same as the separation between the two outer lines of the resolved triplet in the dilute salt, and suggests that we are dealing with an unresolved triplet. This is supported by the fact that the narrowest line is obtained in an intermediate direction corresponding to the angle where the overall separation of the triplet would be a minimum, and here h is of the same order of magnitude as would be expected from magnetic interaction alone. If the observed variations in h were entirely due to an unresolved triplet structure, it would correspond to splitting parameters of the order $D \simeq 0{\cdot}045\ \mathrm{cm}^{-1}$; $E \simeq 0{\cdot}013\ \mathrm{cm}^{-1}$. These are not far from the values expected from extrapolation of the results obtained in the dilute salts, if we assume the variation with diluent to be linear in the atomic number (Mn is one, and Co three units beyond Cr).

If D were as large as this, however, one would expect the line shape to be that of an unresolved triplet, whereas the observed shape (see figure) is too peaked in the centre and too extended in the wings. This suggests that there is also appreciable exchange interaction between neighbouring complexes.



Line shapes in broadest and narrowest directions in an undiluted crystal of $K_3Cr(CN)_6$.

Table 3. Root Mean Square Line Widths in undiluted $K_3Cr(CN)_6$

	a axis	*b* axis	*c* axis
R.M.S. width (kG)	0·75	0·24	0·70
R.M.S. width (cm⁻¹)	0·070	0·022	0·065

5.2. *Potassium Ferricyanide*

The spectrum of undiluted $K_3Fe(CN)_6$ was examined at 20°K by Bleaney and Ingram (1952), but further measurements have revealed that the behaviour is much more complex than was thought. The principal features are summarized below.

(*a*) *Attempts to measure g-values.*

Only one resonance line is observed in any direction and hence the principal axes of its g-tensor should coincide with the crystallographic axes a, b, c. Attempts to measure the g-values yielded variable results, the spread being considerably greater than would be expected from the difficulty in locating the position of maximum absorption on a rather broad line. If this maximum corresponds to the centre of the line, the *g*-value calculated therefrom for a given axis (say the *a* axis) depends on the orientation of the oscillatory magnetic field. This is illustrated in table 4, where the resonance results are given in the first two lines.

Table 4. Apparent *g*-values for $K_3Fe(CN)_6$

	a axis	*b* axis	*c* axis
H_{rf} along *c* axis	2·36	2·21	—
H_{rf} normal to *c* axis	2·31	2·39	0·94
Extrapolated from Jackson (1938)	$2{\cdot}20 \pm 0{\cdot}02$	$2{\cdot}28 \pm 0{\cdot}01$	$0{\cdot}95 \pm 0{\cdot}01$
From susceptibility at He temperatures	$2{\cdot}21 \pm 0{\cdot}04$	$2{\cdot}28 \pm 0{\cdot}04$	$0{\cdot}98 \pm 0{\cdot}02$

The measurements for the *a* and *b* axes were made at about 0·8 cm^{-1}, that for the *c* axis (which needs rather a large magnetic field) at about 0·24 cm^{-1}. A striking feature of the results is that in the *ab* plane, the directions of maximum and minimum *g* values are interchanged according to the orientation of the oscillatory field H_{rf}, even though H_{rf} is always normal to the steady field H_0. A second anomalous effect was obtained by making measurements with H_{rf} in a fixed direction in the *ab* plane, and measuring the value of H_0 at which maximum absorption occurred as the steady field was rotated in the *ab* plane. Using radiation of 0·3 cm^{-1} normal values of H_0, similar to those given in table 4, were obtained when the steady field was normal to H_{rf}, but as the angle between H_0 and H_{rf} diminished the value of H_0 went to abnormally high or low values depending on whether for normal behaviour it should be rising or falling as the angle approached zero. As the magnet was rotated through the zero angle a discontinuity in the value of H_0 for maximum absorption occurred. By repeating the experiment with different orientations of H_{rf} relative to the *a*, *b* axes it was verified that the anomalous behaviour occurred always as the direction of H_0 approached that of H_{rf}.

In view of these anomalies, little significance can be attached to the *g* values derived from the resonance measurements, though these appear to be roughly the same as those calculated from susceptibility measurements. These are also given in table 4: the third set is calculated from the measurements of Jackson (1938) which extended down to 14°K, after allowing for the temperature independent paramagnetism; the fourth set gives the values found by McKim and Wolf (1956) from measurements down to 1°K. The *g*-values are close to those measured along the crystallographic axes in the dilute salt (2·29, 2·17 and $0{\cdot}91_5$ respectively), except that the values for the *a* and *b* axes are interchanged.

(*b*) *Measurement of line width.*

The half width of the line at half intensity was measured by a point-by-point plot of the lines, as described by Bagguley, Bleaney *et al.* (1948). The values obtained lay between 1 and 2 kG, whereas that expected from purely magnetic interaction between the ions is much smaller ($\simeq$0·2 kG). This excess line width is not due to rapid spin–lattice relaxation as it does not decrease on lowering the temperature from 20°K, while it does decrease on diluting the salt with a diamagnetic compound. The excess width must therefore be due to anisotropic exchange interaction (isotropic exchange would narrow the lines, not broaden them).

Theoretically, the quantity of greatest interest is the root mean square line width, and an attempt has been made (Baker and Bleaney 1955) to relate this to the Weiss constant of the powder and hence obtain the magnitude of the exchange interaction. The r.m.s. line widths used were obtained from unpublished work of Bleaney and Ingram, but subsequent measurements have shown that the line width for a given direction depends on the orientation of the oscillatory magnetic field. For instance, values differing by a factor 2 have been obtained for the *b* axis, according to whether H_{rf} was along or perpendicular to the *c* axis.

Theoretical analysis of the effects to be expected in a concentrated salt when anisotropic exchange is present is very complicated, but a simple model of two interacting ions illustrates how some of the observed effects may arise. With effective spin 1/2 for each ion, the energy levels split into a singlet and a triplet, transitions being allowed only within the triplet if the two ions are identical. When the exchange is anisotropic this triplet is split like the ground state of a Ni^{2+} ion in a rhombic field, and the two strongly allowed transitions are displaced above and below the normal position for an unsplit triplet. The relative intensity of these two lines depends on the orientation of the oscillatory field; thus if the steady field is along the z axis, one line will be the stronger when H_{rf} is along the x axis, and the other will be the stronger when it is along the y axis. Also, if the steady and oscillatory magnetic fields are parallel, resonance absorption would be observed at a field largely determined by the splitting of the triplet, that is, by the anisotropic exchange interaction. In a concentrated salt the energy levels will be blurred by interaction with a number of neighbours, but the effects outlined above would still be expected. Thus the line width is increased through the splitting of the two resonance lines, and though these are not resolved, the position of maximum absorption may be shifted and depend on the orientation of H_{rf}. The presence of strong exchange interaction would also explain why two resonance lines due to differently oriented ions were not observed, through a mechanism similar to that proposed for copper sulphate (Bagguley, Griffiths and Pryce 1948).

Though no detailed quantitative conclusions about the exchange interaction can be drawn from these results, it seems that there must be such interaction with considerable anisotropy (of the order $0{\cdot}1\ \mathrm{cm}^{-1}$) to explain the anomalous resonance behaviour. This is supported by the measurements of McKim and Wolf (1956), which show that the anisotropic exchange is an appreciable fraction of the isotropic exchange interaction. Such a result is not surprising, for the following reason. A simple theory of exchange interaction leads to a cosine coupling ('isotropic exchange') between the real electronic spins; the ground state wave functions of the ferricyanide ion (see Bleaney and O'Brien 1956) are a mixture of several orbital wave functions with different orientations of the electron spin, and a simple cosine coupling between these spins leads to an anisotropic exchange interaction when expressed in terms of the 'effective spin' used to describe the doublet ground state. An anisotropic coupling between the real spins, as might arise from 'super-exchange', would of course also lead to an anisotropic coupling between the effective spins.

REFERENCES

Bagguley, D. M. S., Bleaney, B., Griffiths, J. H. E., Penrose, R. P., and Plumpton, B. I., 1948, *Proc. Phys. Soc.*, **61,** 542, 551.

Bagguley, D. M. S., Griffiths, J. H. E., and Pryce, M. H. L., 1948, *Nature, Lond.*, **162,** 538.

Baker, J. M., and Bleaney, B., 1952, *Proc. Phys. Soc.* A, **65,** 952; 1955, *Proceedings of the International Conference on Low Temperatures* (Paris: Institut International du Froid).

Barkhatov, V., 1942, *Acta Phys., Chim., U.R.S.S.*, **16,** 123.

Barkhatov, V., and Zhdanov, H., 1942, *Acta Phys.-Chim., U.R.S.S.*, **16,** 43.

Bleaney, B., 1951, *Phil. Mag.*, **42,** 441.

Bleaney, B., and Ingram, D. J. E., 1952, *Proc. Phys. Soc.* A, **65,** 953.

Bleaney, B., and O'Brien, M. C. M., 1956, *Proc. Phys. Soc.* B, **69,** 1216.

Bleaney, B., and Stevens, K. W. H., 1953, *Rep. Progr. Phys.*, **16,** 108 (London: Physical Society).

Bleaney, B., and Trenam, R. S., 1954, *Proc. Roy. Soc.* A, **223,** 1.

Bowers, K. D., 1952, *Proc. Phys. Soc.* A, **65,** 860.

Bowers, K. D., and Owen, J., 1955, *Rep. Progr. Phys.*, **18,** 304. (London: Physical Society).

Cooke, A. H., and Duffus, H. J., 1955, *Proc. Phys. Soc.* A, **68,** 32.

Grailich, J., and von Lang, V., 1858, *S. B. Akad. Wiss. Wien.*, **32,** 43.

Groth, P., 1906, *Chemische Kristallographie* (Leipzig: Verlag von Wilhelm Engelmann), Vol. 1, 418, 325.

Guha, B. C., 1951, *Proc. Roy. Soc.* A, **206,** 353.

Jackson, L. C., 1933, *Proc. Roy. Soc.* A, **140,** 695; 1938, *Proc. Phys. Soc.*, **50,** 707.

Kotani, M., 1949, *J. Phys. Soc. Japan*, **4,** 293.

Masson, H., 1947, *C. R. Acad. Sci., Paris*, **224,** 1277.

McKim, F. R., and Wolf, W. P., 1956, *Proc. Phys. Soc.* B, **69,** 1231.

Pauling, L., 1931, *J. Amer. Chem. Soc.*, **53,** 1367.

Pospelov, V. A., and Zhdanov, G. S., 1947, *Z. fiz. Khim., S.S.S.R.*, **21,** 405.

Stevens, K. W. H., 1953, *Proc. Roy. Soc.* A, **219,** 542.

Van Vleck, J. H., 1935, *J. Chem. Phys.*, **3,** 807.

Paramagnetic Resonance in some Complex Cyanides of the Iron Group

II. Theory

By B. BLEANEY and M. C. M. O'BRIEN
Clarendon Laboratory, Oxford

MS. received 31*st August* 1955, *and in revised form* 19*th June* 1956

Abstract. The theory of Stevens for covalent XY_6 complexes is developed for the cases of d^3 and d^5 with low symmetry, and compared with the magnetic resonance results of the preceding paper. The effect of the coupling to the doublet states on the spin Hamiltonian constants for the d^3 complexes is found to be small compared with that of coupling to the quartet states, although the latter are probably further removed energetically. In the d^5 complexes it is shown that the relation between the *g*-values and the susceptibility does not depend on the orientation of the principal axes of the *g* tensor relative to the complex; the principal values of the susceptibility of $K_3Fe(CN)_6$ are calculated from the *g*-values and compared with earlier experimental determinations, the value of the spin–orbit coupling being adjusted to give the best fit. The observed hyperfine constants for $K_4Mn(CN)_6 3H_2O$ are compared with the theoretical expressions and can be explained with reasonable values of the parameters. The orbital *g*-factor *k* is found to be 0·87 for the ferricyanide and 0·74 for the manganocyanide. An appendix discusses the conditions under which hyperfine terms of low symmetry can appear in the spin Hamiltonian, and the effect they may have on the observed spectrum.

---

§ 1. Introduction

The complex cyanides of the iron group show a magnetic behaviour that is markedly different from that of the hydrated ions. There are stable hydrated ions with anything from none to ten d-electrons, and the susceptibility and paramagnetic resonance results show that the ground state has the spin and orbital angular momentum predicted by Hund's rule, with the degeneracy in the orbital levels partly lifted by the asymmetry of the surroundings. The cyanides with octahedral complexes $M(CN)_6$ are not stable with more than six d-electrons, and their spins and orbital angular momenta can be predicted on a modified Hund's rule, in which the orbital degeneracy of the d-shell is reduced from five to three. Pauling (1931) proposed that the spin of the ground state of a complex should be taken as the criterion of whether the complex is covalently bonded or not, those complexes behaving like the cyanides being covalent. This was based on the assumption that the bonding energy is always overwhelmingly large compared with the coulomb interactions between the d-electrons of the central ion which tend to cause obedience to Hund's rule, and Van Vleck (1935) showed that the same effect could be produced on a crystal field theory if the field were large enough.

The effect of a perturbation of cubic symmetry, whether treated on a crystal field or a bonding model, is to split the states of a d-electron into two sets, the dϵ and dγ manifolds, and the strength of the perturbation is given by the magnitude of the splitting K. The dϵ and dγ manifolds belong to the two irreducible representations of the cubic group, T_2 and E (Mulliken's notation), and their wave functions have the same angular variations as the expressions

$$\mathrm{d}\epsilon \begin{cases} xy \\ yz, \\ zx \end{cases} \qquad \mathrm{d}\gamma \begin{cases} x^2-y^2 \\ 3z^2-r^2 \end{cases}$$

The dγ wave functions are those with lobes pointing towards the ligands, and have the higher energy on any reasonable bonding or crystal field model. They have the right symmetry to form σ-bonds with the ligands, while the dϵ orbitals can form π-bonds.

If there are several d-electrons the cubic perturbation will have to compete with the coulomb interactions between them; if K is very large the electrons will fill the dϵ levels before going into the dγ levels, while large coulomb interactions will tend to keep all the spins parallel by allowing only one electron in each level. The behaviour of the hydrated ions of the iron group corresponds to K being smaller than the coulomb interactions and that of the complex cyanides to K being larger. The observed spins of the ground states of the cyanides require only that K just predominates over the coulomb interactions, and a detailed investigation of the theory of the paramagnetic resonance spectrum of the chromicyanide ion has been made to find the relative sizes of the two effects.

Stevens (1953) has shown that in the explanation of the properties of an octahedral covalent complex the formation of weak π-bonds between the central ion and the ligands must be considered. The effects of such bonding appear in the matrix elements of the orbital angular momentum, the spin–orbit coupling and the hyperfine interaction. Each orbital in the dϵ manifold is replaced by a linear combination of dϵ orbitals on the central atom and π-orbitals on the ligands, this linear combination having the same transformation properties under the rotations and reflections of the cubic symmetry group.

The matrices of orbital angular momentum and spin–orbit coupling in these molecular orbitals will have the same transformation properties as the matrices in the dϵ orbitals, and hence the matrix elements can conveniently be represented as dϵ matrix elements multiplied by some factor. For instance, in a complex of cubic symmetry, the operator $\mathbf{L}$ operating in the manifold of molecular orbitals transforming like dϵ orbitals can be replaced by the operator $k\mathbf{L}$ operating in the dϵ manifold. The factor k has been called the orbital g factor, and $k \leqslant 1$ for any reasonable value of the overlap between the different atomic orbitals. In a similar way the spin–orbit coupling

$$\mathscr{H} = \sum_j \frac{e}{2m^2c^2}(\operatorname{grad} V) \times \mathbf{p}_j \,.\, \mathbf{s}_j$$

operating in the manifold of molecular orbitals is replaced by $\zeta'\Sigma\mathbf{l}_j\,.\,\mathbf{s}_j$ operating in the dϵ manifold, where ζ' should not be very different from ζ, the spin–orbit coupling constant for the dϵ atomic orbitals. Both k and ζ' will be treated as parameters to be determined, though these substitutions are strictly correct only for cubic symmetry. With lower symmetry more parameters should be introduced, but there is then insufficient experimental evidence to determine them all, and we shall approximate by using the forms appropriate to cubic symmetry.

The π-bonding can have the effect of introducing an additional hyperfine structure into the spectrum if the ligands contain nuclei with magnetic moments (Griffiths, Owen and Ward 1953), but this is not observed in the complex cyanides because carbon and nitrogen have zero or very small nuclear moments. The spreading of the wave function of the magnetic electrons has the effect of reducing the electron density on the central ion, and consequently of reducing the magnitude of the hyperfine structure due to the nucleus of the central ion. The expressions for the hyperfine structure all contain the quantity $\langle r^{-3} \rangle$, the average value of r^{-3} in the radial wave function of the dϵ electrons, a quantity which can only be estimated rather approximately. Hence the reduction may be taken as affecting only $\langle r^{-3} \rangle$, whose value will be regarded as a parameter determined from the experimental data.

The ions in which an orbital g-factor k should be most readily determined are those in which the orbital angular momentum is not completely quenched by the cubic field. Of these those with the configuration dϵ^5 are stable and have a Kramers doublet for their ground state in which resonance can be observed. Observations have been made on two complexes with this configuration, the ferricyanide and manganocyanide complexes. Both these complexes have a very low symmetry, and it will be shown that, as long as the approximation is made that the molecular orbitals are those appropriate to cubic symmetry, the expressions for the principal values of g and the susceptibility do not depend on the orientation of the g-tensor relative to the cubic axes. The expressions for the hyperfine constants, on the other hand, do depend on this orientation, because the interaction between the electron spin magnetic moment and the nuclear magnetic moment is not a function only of the angle between the two moments, but depends also on the distribution of spin density. Calculations have been carried out for some special cases and compared with the experimental results for the manganocyanide.

§ 2. THE THEORY OF THE OCTAHEDRAL d^3 COMPLEX

In a free ion the configuration d^3 contains the quartet terms ^{4}F and ^{4}P and a number of doublet terms; in a cubic field the ground state is always a spin quartet, whatever the size of K, so the perturbations of lowest order in the spin–orbit coupling and magnetic moment will be through the other quartet states, and initially the doublet states will be neglected. The arrangement of the energy levels of the quartet states for various sizes of K and the coulomb interaction is shown in figure 1. An important feature of this scheme is that each of the states $^4A_2$ and $^4T_2$ occurs only once. This means that the specifications of these states as products of single electron states do not depend on the relative sizes of K and the coulomb interactions.

Orbital angular momentum belongs to the irreducible representation T_1 of the cubic group, and since the product $T_1 \times A_2$ transforms like T_2 the angular momentum couples the ground state only to the state $^4T_2$. Since the ground state and the state $^4T_2$ are both derived from the term ^{4}F they both have the same total angular momentum, and the operators $\Sigma_j \mathbf{l}_j$ and $\Sigma_j \mathbf{l}_j . \mathbf{s}_j$ can be replaced by $\mathbf{L}$ and $\mathbf{L} . \mathbf{S}$. With this simplification the terms in the Hamiltonian producing a second-order perturbation in the ground state are $\beta\mathbf{L}$ and $\frac{1}{3}\zeta' \mathbf{L} . \mathbf{S}$, and when these perturbations are put in the form of the usual spin Hamiltonian constants (see equation (1) of the preceding paper) they give:

$$g_\mu = 2{\cdot}002 - \frac{8\zeta' k}{3(K+V_\mu)} \quad (\mu = x, y, z), \qquad \ldots\ldots(1)$$

$$D = \frac{\zeta'}{6k}[g_z - \tfrac{1}{2}(g_x + g_y)], \qquad E = \tfrac{1}{2}\frac{\zeta'}{6k}(g_x - g_y). \qquad \ldots\ldots(2)$$

Here the three parameters V_x, V_y and V_z that have been introduced represen small distortions from the cubic symmetry that split the orbital triplet $^4T_2$ into three states with energies $K+V_\mu$ etc. The orbital *g*-factor *k* is introduced to allow for the effect of charge transfer on the angular momentum, and the effect on the spin–orbit coupling is represented by the use of ζ' instead of ζ. These expressions hold for distortions of any symmetry as long as V_μ/K is small, but if the distortion causes admixing of the state $^4A_2$ with $^4T_1$ or $^4T_2$ the effect of spin–spin interaction (Pryce 1950) may make an appreciable contribution to *D*, *E*.

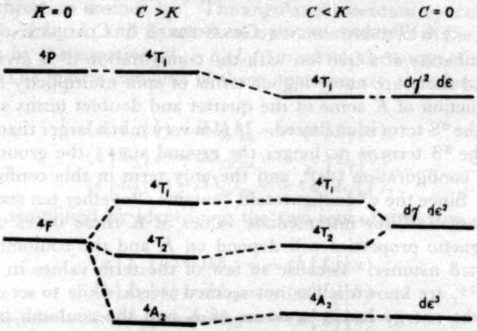


Figure 1. Schematic representation of the energy levels of the quartet states of the configuration d^3 under a perturbation of cubic symmetry. *C* characterizes the coulomb interaction (given by the ^4F–^4P splitting in the free ion); *K* is the cubic splitting.

The effect on the ground state of interaction with the doublet states has been calculated only for the case where *K* is very large, when the configuration is simply $(d\epsilon)^3$. It turns out that the spin Hamiltonian parameters *g*, *D*, *E* are only affected through high order perturbation processes; the change in the value of *g* is of the order 10^{-4}. Also with approximately axial symmetry, the value of *D* is negative. The observed departure of *g* from the free spin value is of order 10^{-2}, and in the chromicyanide a positive value of *D* is found; hence the effect of the quartet states in perturbing the ground state must be more important than that of the doublet states, and the latter will be neglected.

§ 3. COMPARISON WITH EXPERIMENTAL DATA ON d^3

3.1. *The Chromicyanide Complex*

Since *g* is almost isotropic, we consider first the difference of its mean value from the free spin value, which amounts to $0{\cdot}010 \pm 0{\cdot}002$. From equation (1) if V_μ/K is neglected, *k* is taken as 1 and ζ' as 260 cm⁻¹ (the value of ζ for the free ion), the value of K is found to be 70 000 cm^{-1}. This value is about three times as large as that for the hydrated ion. The anisotropy in g calculated from equation (2) is approximately 0·002, which is of the same order as the experimental error.

3.2. *The Vanadocyanide Complex*

The treatment of the vanadocyanide ion follows that of the chromicyanide ion very closely. The spectroscopic data show that in V^{2+} the spin–orbit coupling constant ζ is 155 cm^{-1}, and with the experimental values of the constants in the spin Hamiltonian this leads to $K \simeq 44\,000$ cm^{-1}. This difference in K between isoelectronic divalent and trivalent ions is similar to that found in hydrated ions (Owen 1955).

Owing to the small value of D the anisotropy in g predicted by equation (2) is less than 0·001, which is hardly outside the experimental error in the measurement of g. Hence no detailed check of equation (2) is possible for either d^3 complex.

§ 4. THEORY OF THE OCTAHEDRAL d^5 COMPLEX

The ground state of a free ion with the configuration d^5 is given by Hund's rule as ^{6}S, and there are many higher terms of spin multiplicity four and two. By the introduction of K some of the quartet and doublet terms are lowered in energy while the ^{6}S term is unaltered. If K is very much larger than the coulomb interactions the ^{6}S term is no longer the ground state; the ground term must belong to the configuration (dϵ)^5, and the only term in this configuration is of the type ^{2}T$_2$. Since the d^5 configuration contains altogether ten states belonging to this representation, for intermediate values of K these states will be mixed and their magnetic properties will depend on K and the coulomb interactions in a complicated manner. Because so few of the term values in the 3d^5 ions, Fe^{3+} and Mn^{2+}, are known it has not seemed worth while to set up the secular equation for the energy levels in terms of K and the coulomb integrals; the problem will be treated as if K is overwhelmingly large.

4.1. *The g-values*

By the exclusion of all the configurations except (dϵ)^5 the problem is reduced to that of a single hole in the dϵ shell, which can be treated in exactly the same way as a single electron with a negative spin–orbit coupling coefficient. Both the ferricyanide and manganocyanide complexes have low symmetry in the crystal, so we shall consider the effect of operating on the dϵ orbital wave functions with a perturbation V that is subject to no symmetry restrictions. It will be assumed that the orbitals are those appropriate to cubic symmetry, so that there is a single orbital g-factor k which appears in all the matrix elements of angular momentum.

V will be defined by its eigenvalues and by its eigenstates, which are the following linear combinations of molecular orbitals

$$\begin{aligned} |a\rangle &= l_{zz}|xy\rangle + l_{zx}|yz\rangle + l_{zy}|zx\rangle \\ |b\rangle &= l_{xz}|xy\rangle + l_{xx}|yz\rangle + l_{xy}|zx\rangle \\ |c\rangle &= l_{yz}|xy\rangle + l_{yx}|yz\rangle + l_{yy}|zx\rangle \;. \end{aligned} \qquad \ldots\ldots(3)$$

These states must be orthogonal and normalized, so the set of coefficients, l_{ij}, must behave like the direction cosines of one set of cartesian axes with respect to another set. The states can be labelled in the right order for both sets of axes to be right-handed. Now let L_x', L_y', L_z' be the angular momentum operators with respect to the cubic axes.

Since

$$\langle xy \,|\, L_y' \,|\, yz\rangle = \langle yz \,|\, L_z' \,|\, zx\rangle = \langle zx \,|\, L_x' \,|\, xy\rangle = ik$$
$$\langle yz \,|\, L_y' \,|\, xy\rangle = \langle zx \,|\, L_z' \,|\, yz\rangle = \langle xy \,|\, L_x' \,|\, zx\rangle = -ik$$

and all the other angular momentum matrix elements are zero we have

$$\langle a \,|\, L_j' \,|\, b\rangle = ikl_{yj} = -\langle b \,|\, L_j' \,|\, a\rangle$$
$$\langle b \,|\, L_j' \,|\, c\rangle = ikl_{zj} = -\langle c \,|\, L_j' \,|\, b\rangle$$
$$\langle c \,|\, L_j' \,|\, a\rangle = ikl_{xj} = -\langle a \,|\, L_j' \,|\, c\rangle$$

where j is x, y or z.

The principal axes of the g tensor will make some arbitrary angles with the cubic axes, and the transformation from one set of axes to the other can be made by the set of direction cosines λ_{ij}. The angular momentum operators transform like vectors, so, if L_x, L_y and L_z are these operators referred to the g tensor axes, they are given by the equations $L_j = \lambda_{j\mu} L_\mu'$ (in which the summation convention is assumed). The matrix elements of these operators are, therefore, given by the equations

$$\langle a \,|\, L_j \,|\, b\rangle = ik\lambda_{j\mu} l_{y\mu} = -\langle b \,|\, L_j \,|\, a\rangle$$
$$\langle b \,|\, L_j \,|\, c\rangle = ik\lambda_{j\mu} l_{z\mu} = -\langle c \,|\, L_j \,|\, b\rangle$$
$$\langle c \,|\, L_j \,|\, a\rangle = ik\lambda_{j\mu} l_{x\mu} = -\langle a \,|\, L_j \,|\, c\rangle.$$

These can be simplified by identifying the two sets of direction cosines, when they become

$$\begin{aligned} \langle a \,|\, L_j \,|\, b\rangle &= ik\delta_{jy} = -\langle b \,|\, L_j \,|\, a\rangle \\ \langle b \,|\, L_j \,|\, c\rangle &= ik\delta_{jz} = -\langle c \,|\, L_j \,|\, b\rangle \qquad \ldots\ldots(4) \\ \langle c \,|\, L_j \,|\, a\rangle &= ik\delta_{jx} = -\langle a \,|\, L_j \,|\, c\rangle \end{aligned}$$

where j is x, y or z and δ_{jx} is equal to one when j is x and zero when j is not x. Thus the matrices of the angular momentum operators within the set of eigenstates of V can be made independent of the form of V by a suitable choice of axes.

A unique axis can now be defined by diagonalizing L_z. The appropriate states for this diagonalization are a, $b_+ = (b+ic)/\sqrt{2}$ and $b_- = (b-ic)/\sqrt{2}$. The matrices of the perturbation $\mathscr{H}' = V + \zeta' \mathbf{L}\,.\,\mathbf{S}$ are then:

$$\begin{array}{c|ccc} & a^+ & b_+^{\ -} & b_-^{\ -} \\ \hline a^+ & A & \zeta'/\sqrt{2} & 0 \\ b_+^{\ -} & \zeta'/\sqrt{2} & \frac{1}{2}(B+C)+\frac{1}{2}\zeta' & \frac{1}{2}(B-C) \\ b_-^{\ -} & 0 & \frac{1}{2}(B-C) & \frac{1}{2}(B+C)-\frac{1}{2}\zeta' \end{array}$$

$$\ldots\ldots(5)$$

$$\begin{array}{c|ccc} & a^- & b_-^{\ +} & b_+^{\ +} \\ \hline a^- & A & -\zeta'/\sqrt{2} & 0 \\ b_-^{\ +} & -\zeta'/\sqrt{2} & \frac{1}{2}(B+C)+\frac{1}{2}\zeta' & \frac{1}{2}(B-C) \\ b_+^{\ +} & 0 & \frac{1}{2}(B-C) & \frac{1}{2}(B+C)-\frac{1}{2}\zeta' \end{array}$$

where A, B and C are the eigenvalues of V. All the eigenstates are Kramers doublets, and the ground states can conveniently be written:

$$\left.\begin{aligned} |+m\rangle &= [\sin\alpha\,|a^+\rangle + \cos\alpha\,|b_+^-\rangle]\cos\theta + \sin\theta\,|b_-^-\rangle \\ |-m\rangle &= -[\sin\alpha\,|a^-\rangle - \cos\alpha\,|b_-^+\rangle]\cos\theta + \sin\theta\,|b_+^+\rangle \end{aligned}\right\} \quad \ldots\ldots(6)$$

The g-values are then found by operating within this doublet with the magnetic moment operator, $k\mathbf{L}+2\mathbf{S}$, and this gives:

$$\left.\begin{aligned} g_z &= 2\cos^2\theta\,[\sin^2\alpha - (1+k)\cos^2\alpha] + 2\sin^2\theta\,[k-1] \\ \tfrac{1}{2}(g_x+g_y) &= -2\cos^2\theta\,[\sin^2\alpha + \sqrt{2}k\cos\alpha\sin\alpha] \\ \tfrac{1}{2}(g_x-g_y) &= 2\sin 2\theta\,[\cos\alpha + (k/\sqrt{2})\sin\alpha] \end{aligned}\right\} \quad \ldots\ldots(7)$$

The identification of the states $|\pm m\rangle$ is chosen so that, with $\theta=0$, the matrices of (kL_x+2S_x) and (kL_y+2S_y) in $|\pm m\rangle$ are the same multiple of the Pauli matrices σ_x and σ_y; that is g_x and g_y have the same sign. For a given value of α, this fixes the sign of g_z. Experimentally, the signs of g_x, g_y and g_z are not determined, so two interpretations are possible for a given set, corresponding to alternative sets of states $|\pm m\rangle$ and $|\pm m'\rangle$, which have different values of α, θ, k according as the sign of g_z is assumed to be positive or negative. All the other possible sign differences can be produced by altering the phase factors and identification of the states $|\pm m\rangle$ or $|\pm m'\rangle$. By altering the identification of the different principal axes with the x, y and z axes defined here, solutions for a given set of g values with different values of θ and α can be found that correspond to the same physical ground states; the parameter that does not vary with the choice of axes is k, the value of which can be used to label the two possible different states. Thus for a given set of g values, all the possibilities have been considered when solutions corresponding to two different values of k have been found.

4.2. *The Hyperfine Structure*

The hyperfine structure constants are taken to be those appropriate to the atomic orbitals apart from a reduction in the effective value of $\langle r^{-3}\rangle$, as described in § 1. With this assumption the constants can be calculated by the method of equivalent operators (Abragam and Pryce 1951), the operator $\mathbf{r}$ operating in a manifold of eigenstates of angular momentum being replaced by the operator $\mathbf{L}$. The operators to be used are then

$$N_z = 2\mathscr{P}[L_z + (\tfrac{4}{7}-\kappa)S_z - \tfrac{1}{7}\{L_z(\mathbf{L}\,.\,\mathbf{S}) + (\mathbf{L}\,.\,\mathbf{S})L_z\}]$$

with $\mathscr{P} = 2\gamma\beta\beta_n\langle r^{-3}\rangle$, N_x and N_y having a similar form. κ is the term inserted to allow for small admixtures of s electron wave functions to the ground state.

The matrices of $\mathbf{L}$ and $\mathbf{S}$ in the ground states have already been found, and they have been shown to be independent of the relative orientation of the axes of the g tensor and the cubic axes; the product terms like $L_z(\mathbf{L}\,.\,\mathbf{S})$ involve the matrix elements of $\mathbf{L}$ between dϵ and dγ and these depend on the orientation. Because of this the general case, with arbitrary orientation, has not been considered, but the formulae have been derived for two particular orientations of the principal axis, one or other of which is nearly that found in potassium manganicyanide.

The constants for which formulae will be quoted are those for the following hyperfine interaction:

$$\begin{aligned}&A_xS_xI_x+A_yS_yI_y+A_zS_zI_z+B_{xy}(S_xI_y+S_yI_x)+B_{yz}(S_yI_z+S_zI_y)\\&\quad+B_{zx}(S_zI_x+S_xI_z)+C_{xy}(S_xI_y-S_yI_x)+C_{yz}(S_yI_z-S_zI_y)\\&\quad+C_{zx}(S_zI_x-S_xI_z).\qquad\qquad \ldots\ldots(8)\end{aligned}$$

This is a convenient form of the most general interaction that is linear in **S** and **I** (see Appendix).

With the z axis of the g tensor along a cubic axis, and with the x and y axes each making an angle β with a cubic axis, the constants are given by:

$$\left.\begin{aligned}\frac{1}{2\mathscr{P}}A_z&=\cos^2\theta\left[-\frac{2}{7}-\frac{6}{7}\cos^2\alpha+\frac{3}{7\sqrt{2}}\sin 2\alpha+\frac{\kappa}{2}\cos 2\alpha\right]\\&\qquad+\sin^2\theta\left(\frac{6}{7}+\frac{\kappa}{2}\right)\\\frac{1}{4\mathscr{P}}(A_x+A_y)&=-\cos^2\theta\left[\frac{1}{7}\sin^2\alpha+\frac{11}{14\sqrt{2}}\sin 2\alpha-\frac{\kappa}{2}\sin^2\alpha\right]\\&\qquad-\frac{3}{7}\sin^2\theta\cos 4\beta\\\frac{1}{4\mathscr{P}}(A_x-A_y)&=-\frac{1}{2}\sin 2\theta\left[\frac{1}{7}\cos\alpha-\frac{17}{7\sqrt{2}}\sin\alpha+\kappa\cos\alpha\right]\\\frac{1}{2\mathscr{P}}C_{xy}&=-\frac{3}{7}\sin^2\theta\sin 4\beta\end{aligned}\right\}\ \ldots\ldots(9)$$

All the other hyperfine structure constants are zero.

With the z axis along a trigonal axis of the cubic system, and with the x axis making an angle ϕ with the projection of a cubic axis on the xy plane, the constants are given by:

$$\left.\begin{aligned}\frac{1}{2\mathscr{P}}A_z&=\cos^2\theta\left[\frac{2}{7}\sin^2\alpha-\frac{6}{7}\cos^2\alpha-\frac{1}{7\sqrt{2}}\sin 2\alpha+\frac{\kappa}{2}\cos 2\alpha\right]\\&\qquad+\sin^2\theta\left[\frac{8}{7}+\frac{\kappa}{2}\right]\\\frac{1}{4\mathscr{P}}(A_x+A_y)&=\cos^2\theta\left[\frac{1}{7}-\frac{15}{14\sqrt{2}}\sin 2\alpha+\frac{\kappa}{2}\sin^2\alpha\right]\\\frac{1}{4\mathscr{P}}(A_x-A_y)&=\frac{1}{2}\sin 2\theta\left[-\frac{13}{7\sqrt{2}}\sin\alpha-\frac{1}{7}\cos\alpha+\kappa\cos\alpha\right]\\\frac{1}{2\mathscr{P}}B_{zx}&=-\frac{1}{7}\cos 3\phi[2\sqrt{2}\sin^2\theta+\sin 2\theta\{2\sin\alpha-\sqrt{2}\cos\alpha\}]\\\frac{1}{2\mathscr{P}}B_{yz}&=-\frac{1}{7}\sin 3\phi[2\sqrt{2}\sin^2\theta+\sin 2\theta\{2\sin\alpha-\sqrt{2}\cos\alpha\}]\\\frac{1}{2\mathscr{P}}C_{zx}&=+\frac{1}{7}\cos 3\phi[2\sqrt{2}\sin^2\theta-\sin 2\theta\{2\sin\alpha-\sqrt{2}\cos\alpha\}]\\\frac{1}{2\mathscr{P}}C_{yz}&=-\frac{1}{7}\sin 3\phi[2\sqrt{2}\sin^2\theta-\sin 2\theta\{2\sin\alpha-\sqrt{2}\cos\alpha\}]\end{aligned}\right\}\ \ldots\ldots(10)$$

§ 5. g-VALUES AND SUSCEPTIBILITY OF $K_3Fe(CN)_6$

Since no hyperfine structure is observed in potassium ferricyanide, the data which have to be fitted are the g-values of the ground state and the susceptibility. We begin by demonstrating in a simple manner that the observed g-values cannot be fitted using the strong crystal field model, without any π-bonding. The formulae for the g-values would then be just those obtained by putting the normal orbital g-factor ($k=1$) in the expressions (7) and it is then easy to show that the following relations hold:

$$\frac{\frac{1}{2}(g_x-g_y)}{\sqrt{2}\tan\theta}=\frac{\frac{1}{2}(g_x+g_y)}{-\sin\alpha}=\frac{g_z}{\sin\alpha-\sqrt{2}\cos\alpha}. \qquad \ldots\ldots(11)$$

Hence both the unknown constants θ and α are determined by the ratios of the g-values, and on inserting the observed values for the diluted salt $g_x=2{\cdot}35$, $g_y=-2{\cdot}10$, $g_z=-0{\cdot}915$ one obtains $\cot\alpha=1{\cdot}00$, $\tan\theta=-0{\cdot}028$. However, these values of the constants, which give the right anisotropy, lead to the following g-values: $g_x=-2{\cdot}55$, $g_y=-2{\cdot}28$, $g_z=-1{\cdot}00$, which are all about 8% too large. This is the nearest approach of the theory to the correct values; different values of the parameters can be used to obtain a fit with two of the observed principal g-values, but then a large discrepancy is obtained with the remaining g-values. Again, if we assume that g_z is positive, but the other g-values are negative, and fit the anisotropy, then the calculated g-values are 21% too high.

When a single orbital g-factor k is introduced, so that the theory contains an extra adjustable parameter, the observed g-values of the ground state can be used to determine the three unknowns α, θ and k. If all the g-values are assumed to be negative, these parameters are found to have the values

$$\cot\alpha=1{\cdot}00_6, \qquad \tan\theta=-0{\cdot}028, \qquad k=0{\cdot}87_5$$

so that the values of α and θ are not markedly different from those obtained above, and the reduction in the orbital g-factor from unity is not very great. For the alternative of g_z positive, and the other g-values negative, the values

$$\cot\alpha=0{\cdot}417, \qquad \tan\theta=-0{\cdot}040, \qquad k=0{\cdot}56$$

are required, with a very substantial reduction in the orbital g-factor. While there are no *a priori* reasons for ruling out such a low value of k, it turns out that the susceptibility cannot be fitted at all well, and even an approximate fit requires an excessively low value of the spin–orbit coupling. Because of the low value of k, it also leads to values of the susceptibility at high temperatures which are considerably smaller than those observed. This alternative will therefore be rejected.

In calculating the susceptibility of the concentrated salt we are faced by the difficulty that values of the principal g-values and the directions of the principal axes are known only for the dilute salt. In the concentrated salt only the values of g along the pseudo-orthorhombic axes a, b, c are known; however, for both ions in unit cell the z axis coincides with the c axis, while the a and b axes bisect the angles between the respective x and y axes of the two ions. The observed g-values in the concentrated salt do not differ greatly from those in the dilute salt, and we shall calculate the susceptibility from parameters based on the latter.

Calculation of the susceptibility requires diagonalization of the energy matrices (5), since the ground doublet has a considerable temperature independent paramagnetism and at room temperature two excited doublets have appreciable population. Since θ is very small and α is very close to 45°, the energy levels are found to be $+\zeta'/\sqrt{2}$, $-\zeta'/\sqrt{2}$ and $-\zeta'$, using a good approximation to diagonalize the matrices, where the amount by which the states are not exactly normalized or orthogonal is of the order $\sin^2\theta$ or approximately 10^{-3}.

The susceptibility along a principal axis is conveniently specified by the value of p^2, the square of the Bohr magneton number, defined by the relations

$$\chi_x = \frac{Np_x^2\beta^2}{3kT}, \text{ etc.} \qquad \ldots\ldots(13)$$

The values of p_x^2, etc. are found by standard methods, and may be expressed in the form

$$p_x^2 = \{(4{\cdot}14 + 5{\cdot}39w) + (0{\cdot}028 + 0{\cdot}70w)\gamma - 6{\cdot}09w\delta\}(1+\gamma+\delta)^{-1}$$
$$p_y^2 = \{(3{\cdot}31 + 5{\cdot}97w) + 0{\cdot}048\gamma - 5{\cdot}97w\delta\}(1+\gamma+\delta)^{-1}$$
$$p_z^2 = \{(0{\cdot}628 + 8{\cdot}76w) + (0{\cdot}530 - 8{\cdot}76w)\gamma + 0{\cdot}047\delta\}(1+\gamma+\delta)^{-1}$$

where $w = -kT/\zeta'$, $\gamma = \exp(-\sqrt{2}/w)$, $\delta = \exp[-(1+\sqrt{2})/\sqrt{2}w]$. These formulae refer to the principal axes of the individual ions, and for the pseudo-orthorhombic axes of the crystal

$$p_a^2 = (\tfrac{3}{4}p_x^2 + \tfrac{1}{4}p_y^2), \quad p_b^2 = (\tfrac{1}{4}p_x^2 + \tfrac{3}{4}p_y^2), \quad p_c^2 = p_z^2$$

since the *x* and *y* axes make an angle of 30° with the *a* and *b* axes respectively.

The formulae involve one undetermined parameter, $w = -kT/\zeta'$ where ζ' is the effective value of the spin–orbit coupling for a single hole in the (dϵ) sub-shell. The best fit that can be obtained with the experimental values of the susceptibility is shown in figure 2, where the full lines indicate the calculated variation of p^2 for each of the *a*, *b*, *c* axes, and the experimental points are taken from the data of Jackson (1938) and Masson (1947). Earlier results of Jackson (1933) are in close agreement with these values, but later results of Guha (1951) are considerably (10 to 20%) lower. Guha's work involves separate measurements of the anisotropy and of the absolute susceptibility in one direction only, and the anisotropy observed by him agrees with the other experiments much more closely than the absolute values (the position is rather similar in copper acetate). In view of the excellent agreement between the two quite independent determinations by Jackson and the third by Masson, we have discarded the results of Guha. Figure 2 shows that the calculated results do not anywhere differ from the experimental results by more than a few per cent, if a value of $-400°\text{K} = -278\,\text{cm}^{-1}$ is taken for ζ'. This is considerably smaller than the value of ζ for a free Fe^{3+} ion, which is estimated as about $625°\text{K} = 435\,\text{cm}^{-1}$ (the difference of sign arises because we are dealing with a hole in the (dϵ) shell). Thus the apparent reduction in the spin–orbit coupling ($\zeta'/\zeta = 0{\cdot}64$) is rather greater than that in the orbital *g*-factor ($k = 0{\cdot}87$).

The experimental behaviour of the susceptibility for the *a* and *b* axes is rather unusual, showing a reversal in the sign of the anisotropy in the *ab* plane at about 70°K. We have verified that this reversal is real by suspending a crystal with the *c* axis vertical in a horizontal magnetic field when it was found that the *a* axis sets parallel to the field at 90°K, but at 20°K the *b* axis sets along the field. In a dilute crystal the *a* axis sets parallel to the field at all temperatures. These results suggest that there is a small change in the crystal field with temperature in the concentrated salt; either a rotation of the principal axes of each ion in the *ab* plane, or a reversal in the sign of the rhombic field for each ion, would explain the observed behaviour of the susceptibility. The rhombic anisotropy arises from the non-zero value of $\tan\theta$, but the value of $-0{\cdot}028$ required to fit the observed anisotropy in the dilute salt is so small that a very slight change in the crystal field could reverse the sign of the anisotropy. This would not alter the calculations of the susceptibility materially; the mean of the susceptibility for the x and y axes (which is also the mean for the a and b axes) is practically independent of the value of $\tan\theta$ when it is so small.

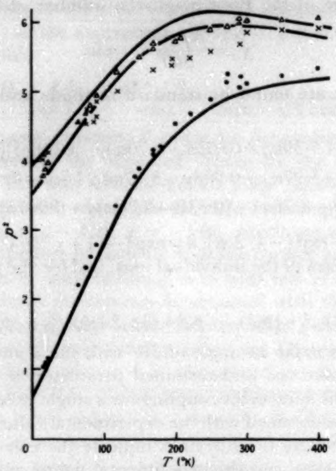


Figure 2. p^2 (square of the Bohr magneton number) for $K_3Fe(CN)_6$ as a function of the temperature. The full curves are calculated from the theory (a, b, c axes in descending order). The points are from experimental data of Jackson, except the three points at 300°, 350° and 400°K, which are from empirical equations of Masson.

§ 6. THE g-VALUES AND HYPERFINE STRUCTURE OF $K_4(Mn, Fe)(CN)_6, 3H_2O$

Potassium manganocyanide is isomorphous with potassium ferrocyanide, whose structure has been partly determined by Pospelov and Zdanov (1947). The symmetry is monoclinic, with the a and c axes equal in length and inclined at an angle of 90° ± 5°. There are four inequivalent cyanide octahedra in the unit cell, but they all have the same orientation, with one axis pointing along the b axis of the crystal. Magnetically there are two inequivalent ions in unit cell, which transform into each other by a reflection in the ac plane. The g-values of these ions, measured on the diluted salt, as reported in the previous paper, are given in the table.

As was explained in § 4, any set of g-values for which the signs are not known can be interpreted in two ways: if two of the g-values are assumed to have the same sign then the set corresponds to different distortions of the complex according as the third g-value is positive or negative. The two interpretations are characterized by different values of k, and for this set of g-values these are 0·93 and 0·74. $k=0{\cdot}74$ corresponds to the greater overall splitting of the orbital levels by the distortion, about 500 cm^{-1}.

Because the angle of the crystal is so close to a right angle it was found to be impossible to distinguish between the a and c axes; hence there is an ambiguity in assigning the magnetic axes to the axes of the octahedron, and the two possible ways of doing so must both be considered. The experimental results were fitted to a spin Hamiltonian of the form $\mathscr{H}=\beta\mathbf{H}\,.\,g\,.\,\mathbf{S}+\mathbf{S}\,.\,A\,.\,\mathbf{I}$ with the assumption that the tensors A and g both have the same principal axes. The principal values of A and g are given in the table below.

axis	g	$A(10^{-4}\text{cm}^{-1})$
(i)	$2{\cdot}624 \pm 0{\cdot}008$	$84{\cdot}5 \pm 0{\cdot}5$
(ii)	$2{\cdot}182 \pm 0{\cdot}008$	$46{\cdot}5 \pm 0{\cdot}5$
(iii)	$0{\cdot}63 \pm 0{\cdot}10$	104 ± 20

With this labelling of the axes of the g tensor their positions relative to the axes of the octahedron can be specified; either the axis (ii) lies about 10° off a trigonal axis, with $3\phi=\pi/2$ if Ox is identified with the axis (i), or the axis (i) lies about 12° off an octahedral axis with $\beta=25°$.

The two possible values of k together with the two possible orientations allow for the construction of four possible sets of hyperfine structure constants. These were calculated with the assumptions that $\cos 4\beta=0$, $3\phi=\pi/2$, and that all the B_{ij} and C_{ij} are negligible. The set of three equations for A_x, A_y and A_z then contains only the two parameters $\mathscr{P}$ and κ, and a graphical method was adopted for finding whether there are any possible solutions. It was found that, with $k=0{\cdot}93$, there is no pair of values of $\mathscr{P}$ and κ with which A_x, A_y and A_z can all be fitted within the experimental error. With $k=0{\cdot}74$ the equations can be solved with either of the orientations, the values of the parameters being:

(1) Tetragonal axis: $\kappa\mathscr{P}/\gamma \doteq 0{\cdot}006_8\,\text{cm}^{-1}$, $\mathscr{P}=0{\cdot}011_5$

(2) Trigonal axis: $\kappa\mathscr{P}/\gamma=0{\cdot}007_8\,\text{cm}^{-1}$, $\mathscr{P}=0{\cdot}013_8$

Here κ is replaced by $\kappa\mathscr{P}/\gamma$ where γ is the nuclear g factor ($\gamma=1{\cdot}38$ for ^{55}Mn).

The quantity $\kappa\mathscr{P}/\gamma$ has been found to vary very little between the ions in the 3d group that have been studied by the method of paramagnetic resonance (see Bleaney and Stevens 1953) and it probably does not depend much on the surroundings of the ion. The average value is 0·006 cm^{-1} for ions surrounded by water molecules, and it appears to be rather less in the vanadocyanide and chromicyanide complexes, in which almost the whole hyperfine structure is due to this term. The quantity $\langle r^{-3}\rangle$ can be estimated from the spin–orbit coupling constant by Goudsmit's formula, as is usually done in optical spectroscopy. The spin–orbit coupling constant for Mn^{2+} has been estimated as 300 cm^{-1} by interpolation in the isoelectronic series and in the series of manganese ions with different degrees of ionization; the value quoted by Kotani (1949) appears to be in error. With this value of the spin–orbit coupling we find that $2\gamma\beta\beta_n\langle r^{-3}\rangle=0{\cdot}016\ \text{cm}^{-1}$. $\mathscr{P}$ would be less than this by a normalizing factor of which the best estimate is probably to put it equal to k. Thus both pairs of values of $\kappa\mathscr{P}/\gamma$ and $\mathscr{P}$ are quite reasonable.

An attempt has been made to find out whether the experimental results enable one to distinguish between the near-tetragonal and near-trigonal distortions. Using the values of θ, α and $\mathscr{P}$ obtained above, the extra terms in the hyperfine interaction tensor (8) were calculated. In the 'tetragonal' case (equation (9)) C_{xy} is found to be 0·001 cm^{-1}, and in the other case (equation (10)) $B_{yz}=0{\cdot}003$ cm^{-1}, $C_{yz}=0{\cdot}0005$ cm^{-1}, $B_{zx}=C_{zx}=0$. Using these values a theoretical plot of the quantity R^2g^2 (see Appendix) against $l_y{}^2$ was made in each case, and compared with an experimental determination of these quantities. Unfortunately, the latter extended only over part of the range as it was not possible to make measurements in directions where the g-value was very low. Both the theoretical plots (which are based on the extrapolated value of A for axis (iii)) deviate little from the straight line through the experimental points. The scatter on these points is such that the 'trigonal' case (which has the greater curvature) cannot be excluded, though the fit with the 'tetragonal' case is appreciably better.

APPENDIX

The Magnetic Hyperfine Structure in a Spin Hamiltonian of low Symmetry

In calculating the hyperfine interaction in the manganocyanide complex it was found necessary, because of the low symmetry, to introduce into the spin Hamiltonian some terms of the form $S_\lambda I_\mu$. The effects of such terms and the conditions under which they are allowed will be considered in this section.

The general spin Hamiltonian for a spin of a half, neglecting the quadrupole term and the direct effect of the field on the nucleus, is $\mathscr{H}=\beta\mathbf{H}\,.\,g\,.\,\mathbf{S}+\mathbf{S}\,.\,A\,.\,\mathbf{I}$. Here g and A are tensors, $\mathbf{H}$ is a vector and $\mathbf{S}$ and $\mathbf{I}$ are operators that transform like vectors. It is convenient to write this in the form $\mathscr{H}=\beta g_{\lambda\mu}H_\lambda S_\mu+A_{\lambda\mu}S_\lambda I_\mu$ with the use of the summation convention. $g_{\lambda\mu}$ is a symmetric tensor, and the Hamiltonian is usually referred to its principal axes thus:

$$\mathscr{H}=\beta\sum_k g_k H_k S_k+A_{\lambda\mu}S_\lambda I_\mu.$$

If $\mathbf{S}$ were the operator representing a real spin, $\mathbf{S}$ and $\mathbf{I}$ would commute, and A would be a symmetric tensor. If $\mathbf{S}$ represents an effective spin the term $A_{\lambda\mu}S_\lambda I_\mu$ is replacing another of the form $\mathbf{N}\,.\,\mathbf{I}$ where $\mathbf{N}$ is the hyperfine structure operator (Abragam and Pryce 1951)

$$\mathbf{N}=2\gamma\beta\beta_n\sum_k\left[\frac{\mathbf{l}_k-\mathbf{s}_k}{r_k{}^3}+\frac{3(\mathbf{r}_k\,.\,\mathbf{s}_k)\mathbf{r}_k}{r_k{}^5}+\frac{8\pi}{3}\delta(\mathbf{r}_k)\mathbf{s}_k\right].$$

The matrix elements of this operator in the ground doublet are not necessarily multiples of the matrix elements of real spin, but any 2 × 2 matrix can be expressed as a linear combination of the Pauli matrices, so we may write $N_\mu=A_{\lambda\mu}S_\lambda$. $A_{\lambda\mu}$ is a tensor and it is sometimes convenient to write it as the sum of its symmetric and antisymmetric parts: $A_{\lambda\mu}=B_{\lambda\mu}+C_{\lambda\mu}$ with $B_{\lambda\mu}=B_{\mu\lambda}$ and $C_{\lambda\mu}=-C_{\mu\lambda}$.

The spin Hamiltonian must conform to the total symmetry about the ion in the crystal, and if the symmetry is high the number of independent constants may be very small. Under the rotations and reflections of the point groups symmetric tensors transform like the coefficients of a central quadric. Because of this $B_{\lambda\mu}$ and $g_{\lambda\mu}$ are different from zero only if there is a linear combination containing $x_\lambda x_\mu$ which is invariant under all the transformations of the symmetry group. The invariance of a form such as x^2+y^2 requires that $B_{xx}=B_{yy}$ and $g_{xx}=g_{yy}$, while if there is no combination of xy, yz and zx that is invariant the principal axes of both tensors are coincident, and their positions are determined by the symmetry. An antisymmetrical tensor transforms like a vector product; the coefficient $C_{\lambda\mu}$ is different from zero only if there is an invariant linear combination containing $\epsilon_{\lambda\mu\nu}x_\nu$. $\epsilon_{\lambda\mu\nu}$ is zero unless λ, μ and ν are all different, $+1$ if they are in the same cyclic order as the coordinate axes taken right-handedly, and -1 if they are in the opposite order; it is invariant under rotation, and changes sign under reflection and inversion.

Using this method it can be shown that under cubic symmetry the spin Hamiltonian must take the form with complete rotational symmetry: $\mathscr{H}=g\beta\mathbf{H}.\mathbf{S}+A\mathbf{S}.\mathbf{I}$. Under any group containing an axis of threefold or higher symmetry the spin Hamiltonian takes the axial form that is usually written

$$\mathscr{H}=g_{\parallel}\beta H_zS_z+g_{\perp}\beta(H_xS_x+H_yS_y)+A_{\parallel}S_zI_z+A_{\perp}(S_xI_x+S_yI_y)+C(S_xI_y-S_yI_x)$$

where C is zero if there is inversion symmetry or a plane of reflection containing the axis.

It is of interest to see what the energy levels of this general Hamiltonian are like. The magnetic field along a general direction can be represented as: $H_\lambda=l_\lambda H$, $l_\lambda l_\lambda=1$. If the term [illegible] .**S** is diagonalized the eigenstates can be represented as $|+\rangle$ and $|-\rangle$ an[illegible] the hyperfine splitting is small compared with the difference between $|+\rangle$ and $|-\rangle$ only the terms diagonal in these states are important. These are:

$$\langle\pm|\mathscr{H}|\pm\rangle=\pm\tfrac{1}{2}g\beta H\pm\frac{1}{2g}\sum_{\lambda,\mu}[l_\lambda g_\lambda A_{\lambda\mu}I_\mu]$$

where $g=\{l_\lambda^2g_\lambda^2\}^{1/2}$. By a suitable choice of axes for I this can [illegible] put in the form $\langle\pm|\mathscr{H}|\pm\rangle=\pm\tfrac{1}{2}g\beta H\pm\tfrac{1}{2}RI_z$ with $R^2g^2=\sum\limits_{\lambda,\mu,\nu}g_\lambda g_\mu A_{\lambda\nu}A_{\mu\nu}l_\lambda l_\mu$. [illegible] thus see that, at least to first order in A, the appearance of the spectrum with a strong field in any direction is not altered by the presence of cross terms in the spin Hamiltonian. The spectrum consists of $2I+1$ equally spaced levels, of overall separation R, and it is only the variation of R with direction that depends on which of the $A_{\lambda\mu}$ are different from zero.

In the spin Hamiltonian for a great many ions th[illegible]ipal axes of the tensors g and A coincide, so that $A_{\lambda\mu}=\delta_{\lambda\mu}A_\lambda$ say; t[illegible] $=g_\mu^2l_\mu^2$. Thus a plot of R^2g^2 against one of the squared direction cosines [illegible] in any plane containing the axis μ is a straight line, as originally shown by Pryce (1949), and such a plot is often made in order to extrapolate the hyperfine constants to axes along which resonance cannot be observed. Such a plot would clearly not give a straight line if coefficients of the type A_{xy} are appreciable.

REFERENCES

Abragam, A., and Pryce, M. H. L., 1951, *Proc. Roy. Soc.* A, **205**, 135.

Bleaney, B., and Stevens, K. W. H., 1953, *Rep. Progr. Phys.*, **16**, 108 (London: Physical Society).

Griffiths, J. H. E., Owen, J., and Ward, I. M., 1953, *Proc. Roy. Soc.* A, **219**, 526.

Guha, B. C., 1951, *Proc. Roy. Soc.* A, **206**, 353.

Jackson, L. C., 1933, *Proc. Roy. Soc.* A, **140**, 695; 1938, *Proc. Phys. Soc.*, **50**, 707.

Kotani, M., 1949, *J. Phys. Soc. Japan*, **4**, 293.

Masson, H., 1947, *C. R. Acad. Sci., Paris*, **224**, 1277.
Owen, J., 1955, *Proc. Roy. Soc.* A, **227**, 183.
Pauling, L., 1931, *J. Amer. Chem. Soc.*, **53**, 1367.
Pospelov, V. A., and Zdanov, G. S., 1947, *Z. fiz. khim. SSSR.*, **21**, 405, *Structure Reports*, **11**, 421.
Pryce, M. H. L., 1949, *Nature, Lond.*, **164**, 117; 1950, *Phys. Rev.*, **80**, 1107.
Stevens, K. W. H., 1953, *Proc. Roy. Soc.* A, **219**, 542.
Van Vleck, J. H., 1935, *J. Chem. Phys.*, **3**, 807.

The Susceptibility and Magnetic Specific Heat of Potassium Ferricyanide at Low Temperatures

By F. R. McKIM and W. P. WOLF
The Clarendon Laboratory, Oxford University

Communicated by B. Bleaney; MS. received 19*th June* 1956

Abstract. The temperature variation of the susceptibility of single crystals of potassium ferricyanide has been measured between 20° and 1°K, using an a.c. mutual inductance method. The results of measurements along each of the three crystalline axes were fitted to expressions of the form

$$\chi_x = \frac{\lambda_x}{T}\left(1+\frac{\Delta_x}{T}+\dots\right)$$

and values of λ_x and Δ_x deduced. It was found that the effect of higher order terms was very small. The magnetic specific heat was measured at 1·05°K using the relaxation method.

An attempt was made to relate the values of Δ to the measured specific heat and paramagnetic resonance line width, on the basis of exchange interaction with n equivalent nearest neighbours. It was found, however, that only approximate agreement could be obtained, and this with the rather small value of $n=2$.

---

§ 1. INTRODUCTION

RECENT experiments on the line shape obtained in the paramagnetic resonance spectrum of potassium ferricyanide have shown that there is considerable exchange interaction between the paramagnetic ions (Bleaney and Ingram 1952, Baker, Bleaney and Bowers 1956). The purpose of the measurements described here was to investigate the effect of this interaction on the susceptibility, and to see if there was a simple relationship between the variation of susceptibility with temperature, the magnetic specific heat, and the paramagnetic resonance line width.

§ 2. THE MEASUREMENTS

Potassium ferricyanide grows in monoclinic, pseudo-orthorhombic crystals, whose morphology has been described by Groth (1906). Measurements of the susceptibility were made along the three principal axes, using samples cut from the same single crystal, and also on a finely crushed powder. The apparatus and method used have been described elsewhere (Benzie and Cooke 1950). Measurements were made of the mutual inductance M of a coil system surrounding the specimen and the results plotted as a function of $1/T$. In all cases smooth curves were obtained, which tended to straight lines at high temperatures. By extrapolating these lines, the value of M corresponding to $1/T=0$, M_0 say, could be found for each experiment. This quantity, which represented the inductance of the empty coil plus the effect of temperature independent contributions from the specimen, was then subtracted from each measured value of M. The differences $M-M_0$ obtained in this way were directly proportional to the temperature dependent part of the susceptibility of the specimen χ_T, the constant of proportionality depending only on the characteristics of the coil. The value of this constant was determined in a subsidiary experiment using a sphere of chromic alum.

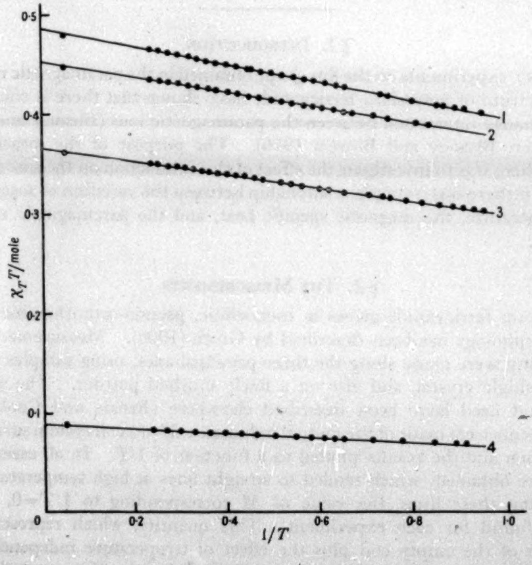
Measurements of the magnetic specific heat were made by the paramagnetic relaxation method due to Casimir and du Pré (1938). The same apparatus was used as that for the previous susceptibility measurements. The method consists in measuring the isothermal susceptibility χ_T and the adiabatic susceptibility χ_S in a steady field of a few hundred gauss. The magnetic specific heat in zero field can then be calculated from these quantities by means of a simple thermodynamic formula.

§ 3. RESULTS

In order to estimate the effect of the exchange interaction on the susceptibility the results were fitted to expressions of the form

$$(\chi_T)_x = \frac{\lambda_x}{T}\left(1+\frac{\Delta_x}{T}+\frac{A_x}{T^2}+\dots\right) \qquad \dots\dots(1)$$

where x denotes one of the principal axes of the crystal. To determine the coefficients in this series the product $(\chi_T)_x T$ was evaluated for each set of measurements and the results plotted as a function of $1/T$. The four graphs obtained are shown in the figure.



The susceptibility of potassium ferricyanide. 1 : *b* axis; 2 : *a* axis; 3 : powdered sample; 4 : *c* axis.

For the measurements along both the a and b axes this plot yielded a straight line within the limits of accuracy of measurement. This implies that, for the susceptibility in these directions, the third and higher order terms in (1) are small at temperatures above 1°K. From the intercepts and slopes of the two straight lines, values of λ_a, λ_b, Δ_a and Δ_b were obtained. Along the c axis small deviations from a straight line were observed at the lowest temperatures. These were interpreted as due to the term in A_c, and an approximate value for A_c was determined. The measurements on a powder again gave a good straight line. This was not inconsistent with the c axis measurement, for the deviation in that direction was small, and as the value of λ_c was also small, in an average for the three directions the deviation would be beyond the limit of detection.

The values of λ and Δ obtained are given in table 1. From the results in the three principal directions the corresponding values for the powder could be computed using the relations

$$\lambda_{\text{powder}} = \tfrac{1}{3}(\lambda_a + \lambda_b + \lambda_c), \quad \Delta_{\text{powder}} = \frac{\lambda_a\Delta_a + \lambda_b\Delta_b + \lambda_c\Delta_c}{\lambda_a + \lambda_b + \lambda_c}.$$

These are also included in the table for comparison with the directly measured quantities. The agreement is very satisfactory.

In order to measure the specific heat by the relaxation method it is necessary to find the adiabatic susceptibility χ_S using a measuring frequency whose period is less than the spin lattice relaxation time τ_L at the particular temperature. For this substance it was found, however, that τ_L was so short at all but the lowest temperatures, that with the highest available frequency (900 c/s), χ_S could only be determined at 1·05°K, and hence only one specific heat measurement could be made. The value of the specific heat constant $b = CT^2/R$ obtained was $b = 0{\cdot}04 \pm 0{\cdot}01$. For the greatest accuracy of measurement, the applied, steady field must be such that the ratio $(\chi_T - \chi_S)/\chi_S$ is approximately unity (Benzie and Cooke 1950). The field required for this depends on the quantity CT^2/λ, which for potassium ferricyanide is so large that even with the strongest field available (~1 kG) the ratio $(\chi_T - \chi_S)/\chi_S$ was only about $\frac{1}{4}$. This accounts for a rather large error in the measurement. In calculating the value of b allowance was made for the deviation of the susceptibility from that for an ideal paramagnetic, using the experimentally determined coefficients in equation (1).

Table 1

	λ (per mole)	λ (Jackson)	Δ (°K)	A (°K)^2
a axis	$0{\cdot}460 \pm 0{\cdot}02$	0·455	$-(0{\cdot}175 \pm 0{\cdot}02)$	$<0{\cdot}01$
b axis	$0{\cdot}490 \pm 0{\cdot}02$	0·490	$-(0{\cdot}190 \pm 0{\cdot}02)$	$<0{\cdot}01$
c axis	$0{\cdot}088 \pm 0{\cdot}004$	0·0845	$-(0{\cdot}255 \pm 0{\cdot}02)$	$+0{\cdot}03 \pm 0{\cdot}01$
Powder measured	$0{\cdot}370 \pm 0{\cdot}02$	—	$-(0{\cdot}200 \pm 0{\cdot}02)$	$<0{\cdot}01$
Powder calculated	$0{\cdot}350 \pm 0{\cdot}02$	—	$-(0{\cdot}190 \pm 0{\cdot}02)$	$<0{\cdot}01$

§ 4. DISCUSSION

We can compare our values of λ with those obtained by extrapolating the measurements made by Jackson (1938) in the range 20° to 14°K. At such temperatures the effects of both exchange interaction and excited states are small, and a good estimate of the Curie constants of the ground doublet can be obtained by extrapolating his values of χT to $T = 0$°K. The values obtained in this way are shown in table 1, and it will be seen that they are in excellent agreement with our results.

In order to compare the observed values of Δ with the results of line width and specific heat measurements, we must consider in more detail the interactions between the different magnetic ions in the crystal. For the simplest case of isolated pairs of identical ions, with effective spins S and $S'=\frac{1}{2}$, we can write the interaction Hamiltonian in the form $\mathscr{H}_{\text{int}}=\mathbf{S}K\mathbf{S}'$ (Kramers 1930) where K is a tensor which includes both magnetic and exchange energies. Referred to the principal axes of this tensor, the spin Hamiltonian in a magnetic field H is then

$$\mathscr{H}=\beta\mathbf{H}g(\mathbf{S}+\mathbf{S}')+K_xS_xS_x'+K_yS_yS_y'+K_zS_zS_z'$$

(Baker and Bleaney 1955). In a concentrated crystal we must consider the effect of several neighbours, the number depending on the variation of K with the distance between the ions.

For this salt it is known from x-ray data (Pospelov and Zhdanov 1947) that there are two types of inequivalent ions per unit cell. On the other hand, paramagnetic resonance measurements have indicated that the two kinds are almost identical magnetically, since all ions have one axis in common with the c axis of the crystal, and only a small anisotropy in the plane at right angles. If we neglect this anistropy and assume that the magnetic axes of all the ions coincide with those of the crystal, we can write down simple expressions for the Weiss constants, specific heat and mean square line widths in terms of the tensors characterizing the interactions between isolated pairs of ions, using the methods developed by Van Vleck (1937, 1948). If there are n_i similar ions for which the interaction with a given ion can be represented by a tensor with components $K_{\alpha\beta}^{(i)}$, then, in a direction x which is one of the magnetic axes of the crystal, the susceptibility is given to a second approximation by

$$(\chi_T)_x=\frac{\lambda_x}{T}\left[1-\frac{1}{4kT}\sum_i n_iK_{xx}^{(i)}+O\left(\frac{1}{T^2}\right)\right]=\frac{\lambda_x}{T}\left[1+\frac{\Delta_x}{T}\right].$$

The expressions for higher order terms in $1/T$ are extremely complicated (Daniels 1953) but it is reasonable to assume that the series converges rapidly at temperatures which are high compared with Δ_x.

The interactions also give rise to a specific heat anomaly, which at high temperatures ($kT\gg K$) is proportional to $1/T^2$. In this region the specific heat is given by

$$\frac{CT^2}{R}=\frac{1}{32k^2}\sum_i\sum_{x,\,y=a,\,b,\,c}n_i(K_{xy}^{(i)}K_{yx}^{(i)}).$$

The effect of the interactions on the mean square width h_x^2 of the line observed in paramagnetic resonance is given in terms of the K's by the formula

$$h_x^2=\frac{9}{16}\sum_i n_i[K_{xx}^{(i)}-\tfrac{1}{3}(K_{xx}^{(i)}+K_{yy}^{(i)}+K_{zz}^{(i)})]^2.$$

Simplified forms of these expressions were given by Baker and Bleaney (1955) who made the assumption that only the interaction with the nearest neighbours was appreciable, and that the axes of K coincided with those of the ions. Under these conditions it is possible to deduce the principal values of K from the measurements of Δ_a, Δ_b and Δ_c if the number of nearest neighbours n is known. The values of CT^2/R and h_a etc. can then be calculated and compared with the measured values. The results of the calculation for $n=2, 4, 6$ are shown in table 2.

Recent work (Baker, Bleaney and Bowers 1956) has shown that the line widths obtained depend on both the direction and frequency of the oscillatory magnetic field used to induce the transitions. For this reason only approximate values are given in the table. It will be seen, however, that for none of the values of n is there agreement between the calculated and measured line widths. For CT^2/R the disagreement is least if we take $n=2$, but this seems rather unlikely in view of the fact that there are six ferric ions almost equidistant from any one site. It is clear, therefore, that the simplifying assumption is not justified, and either the axes of the different K tensors are not the same, or interactions between other than the nearest neighbours are appreciable.

Table 2

	Calculated			Measured
n	2	4	6	
K_a (cm^{-1})	0·24	0·12	0·08	—
h_a (cm^{-1})	0·047	0·033	0·028	~0·1
K_b (cm^{-1})	0·26	0·13	0·09	—
h_b (cm^{-1})	0·026	0·018	0·015	~0·1
K_c (cm^{-1})	0·36	0·18	0·12	—
h_c (cm^{-1})	0·073	0·051	0·042	~0·1
$b=CT^2/R$	0·033	0·0165	0·011	0·04 ± 0·01

In the more general case the measurable quantities CT^2/R, χ_T and h_a etc. are independent, and values for n_i and $K^{(i)}$ cannot be deduced from these measurements alone. The only way of proceeding with the problem would be to obtain further information about the different interaction tensors, by observing paramagnetic resonance in a specimen in which some of the magnetic ions had been replaced by diamagnetic ones. In such a semi-dilute specimen it is sometimes possible to resolve the spectra arising from the interaction between different pairs of ions, and from these it is possible to deduce the principal axes as well as some of the components of the K tensors. Measurements of this type have been made on other substances (Baker and Bleaney 1955, Griffiths and Partridge, private communication).

It is evident from the work described in this paper that there is no simple explanation of the interaction effects in concentrated potassium ferricyanide. The values of Δ_a, Δ_b, Δ_c etc. and CT^2/R have been measured, but the interpretation of the results must await further resonance experiments. Preliminary measurements on semi-dilute crystals have shown that the spectrum contains a very large number of lines, but these have not yet been analysed.

ACKNOWLEDGMENTS

We are indebted to Dr. B. Bleaney and Dr. A. H. Cooke for their interest and for many helpful discussions, to Dr. J. M. Baker for supplying the crystals and to Mr. D. T. Edmonds for his assistance in some of the experiments.

One of us (F. R. M.) is grateful to the President and Fellows of Magdalen College for the award of a Senior MacKinnon Scholarship.

REFERENCES

BAKER, J. M., and BLEANEY, B., 1955, *Conférence de Physique des Basses Températures, Paris.*

BAKER, J. M., BLEANEY, B., and BOWERS, K. D., 1956, *Proc. Phys. Soc.* B, **69, 1205**.

BENZIE, R. J., and COOKE, A. H., 1950, *Proc. Phys. Soc. A*, **63,** 201.

BLEANEY, B., and INGRAM, D. J. E., 1952, *Proc. Phys. Soc.* A, **65,** 953.

CASIMIR, H. B. G., and DU PRÉ, F. K., 1938, *Physica*, **5,** 507.

DANIELS, J. M., 1953, *Proc. Phys. Soc.* A, **66,** 673.

GROTH, P., 1906, *Chemische Krystallographie* (Leipzig: Verlag von Wilhelm Engelmann) Vol. 1, **418,** 325.

JACKSON, L. C., 1938, *Proc. Phys. Soc.*, **50,** 707.

KRAMERS, H. A., 1930, *Proc. Acad. Sci. Amst.* **33,** 959.

POSPELOV, V. A., and ZHDANOV, G. S., 1947, *J. Phys. Chem. Moscou*, **21,** 405.

VAN VLECK, J. H., 1937, *J. Chem. Phys.*, **5,** 320; 1948, *Phys. Rev.*, **74,** 1168.

An Experimental Investigation of Characteristic Electron Energy Losses in Solids

By G. W. JULL†

Imperial College, London

Communicated by D. Gabor ; MS. received 13th July 1956

Abstract. The characteristic electron energy losses in a number of materials have been studied experimentally, with a view to elucidating their origin. The sharp losses of 14·6 ev in Al and 10·0 ev in Mg do not appear in their oxide spectra, while the broad bands in Be and Si are similar to those of their oxides. The mean free path for the 14·6 ev loss process in Al was found to be 260 Å for 10 kev electrons. This strong line is weakened by preparing the film in a number of thin layers, of 20–50 Å each, isolated from one another by some foreign material. This observation suggests that the 14·6 ev loss may well be due to collective excitation, and that boundary layers play an important role in determining the intensity of characteristic losses.

§ 1. Introduction

Characteristic energy losses of medium energy electrons in thin films have been the subject of many investigations in recent years. This interest has been largely stimulated by the suggestion of Bohm and Pines (1952) that excitation of collective oscillations could account for the discrete losses suffered by kilovolt-energy electrons in aluminium and beryllium, as first observed by Ruthemann (1941). The calculated losses were in good agreement with those observed at 14·7 ev in Al and at 18·9 ev in Be, assuming all the valence electrons were free. Moreover, the mean free path calculated by Bohm and Pines for this process was in agreement with the value estimated by Lang (1948) for 7·6 kev electrons in Al. These early transmission measurements have been extended by Möllenstedt (1949), Marton and Leder (1954) and Watanabe (1954) to studies of the losses of 25–40 kev electrons in a wide range of solids. In general their observations show that the loss spectra for metals are more complex than found by Ruthemann and Lang. For most metals several loss-lines were observed, and it is difficult to decide whether collective excitation or single electron interband transitions is the correct interpretation for a particular loss. It has been suggested by Pines (1955), Gabor (1956) and Ferrell (1956) that at least one line in the energy loss spectrum of each metal may be due to collective excitation. Wolff (1953) has noted that in some metals interband transitions will broaden and displace the observed collective loss from its free electron value. But insufficient knowledge of the band structure for most metals has delayed a detailed theory of the role played by interband transitions in determining the characteristic loss spectra—either as the dominant loss mechanism or by modifying collective losses.

While for the most part the experiments do not provide evidence supporting either the single electron excitation or collective interpretation, several observations appear to favour one rather than the other. Marton and Leder (1954) have noted the similarity between the loss spectrum of certain metals and their compounds, and suggested single electron excitation as the loss mechanism in these metals. On the other hand a collective interpretation is favoured by several recent observations. The most clear-cut is the observation by Watanabe (1956) on the angular dependence of energy losses in certain metals. This result appears to give experimental confirmation of the plasma wave dispersion law for losses in Al, Mg and Be. The second by Gabor and Jull (1955) is the observation of a non-linear increase in loss intensity with film thickness for certain losses in gold. Such an effect, which cannot be explained on a single electron excitation basis, appears to indicate that the scattering cross section of collective modes in thin films (less than about 30 Å thickness) is far less than might be expected from observations in thicker films.

---

† Now with the Defence Research Board, Ottawa, Canada.

One striking feature of the experiments, which has made interpretation more difficult, is the disagreement between various workers. For some metals no two investigators find the same spectra. Several likely causes for these discrepancies are the effect of surface contamination, and the fact that many of the losses are rather weak and difficult to detect. Usually films have been prepared by evaporation, and later transferred to the spectrograph for analysis, a procedure which will result in specimen surface contamination, particularly in the case of easily oxidized metals. And for the majority of investigations the recording of the spectra has been done photographically, which makes it difficult to determine absolute loss intensities.

We have attempted to overcome these error sources by constructing an apparatus in which specimen films can be prepared and analysed in the same vacuum, and the spectra then recorded electrometrically. The analyser used was a new type of helical electrostatic spectrometer which was well suited for such energy loss investigation and will be described in detail elsewhere. Using 8–12 kev primary electrons, measurements have been made on the losses in a number of solids, many of which were previously investigated only at higher primary energies of 25–40 kev. Some properties of these losses have been further studied in an attempt to determine their origin. The effect of chemical combination for metals showing either sharp or broad main losses in their spectra has been studied. An investigation was made on the losses observed in aluminium for various film thicknesses, with particular reference to the intensity of the 14·6 ev loss. These experiments are of the same type previously reported for gold, and the results are similar. They are discussed below with reference to the recent theory of Gabor (1956) on collective oscillation in thin metal films.

§ 2. APPARATUS

The velocity analyser used was an electrostatic spectrometer in which the electrons follow helical trajectories. It was first described by Gabor (1951) in his investigation of the imaging properties of an electron beam which follows a helix under the influence of an electrostatic field. The simplest form of field which will result in helical electron trajectories is a cylindrical one

$$\phi = \phi_0 \left(1 - \ln \frac{r}{r_0}\right) \qquad \ldots\ldots(1)$$

where ϕ_0 is the potential on the helical axis of radius r_0. By choosing the diametral pitch of the helix (linear pitch divided by 2π) equal to the radius r_0, Gabor has shown that gaussian image-formation can be obtained. A trajectory starting from a point on the helical axis will return to it after a rotation of 254° ($\pi\sqrt{2}$) around the central axis. It has also been shown that the dispersion Z of the beam in the analyser is

$$Z=\sqrt{2}\pi\epsilon r_0 \qquad \ldots\ldots(2)$$

where ϵ is the energy loss fraction.

For this system the focusing is two-dimensional while the chromatic shift is in the direction of the axis. This may be compared with the case of launching the beam into the field at right angles to the cylindrical axis, for which the focusing is in one direction only, and electrons of different initial energy are spread out radially.

The spectrometer is illustrated in figure 1. A finely collimated electron beam of 8–12 kev energy, incident on the specimen film, was subsequently decelerated to 1·6 kev standard energy before entering the analyser. The beam followed a helical trajectory around the central axis and after a revolution around the axis of 254° fell on a screened collector lying along a meridian plane. The collected currents were amplified using an electrometer valve capable of measuring currents down to 10^{-14} amp. With entrance and collector slit widths of $25\,\mu$, and a radius r_0 of 1·74 cm, an energy resolution of 1·1 ev was obtained with 1·6 kev standard energy electrons. This was sufficient to resolve the fine structure in the energy loss spectrum of all metals. The useful angular aperture of the spectrometer was of the order of 5×10^{-3}.

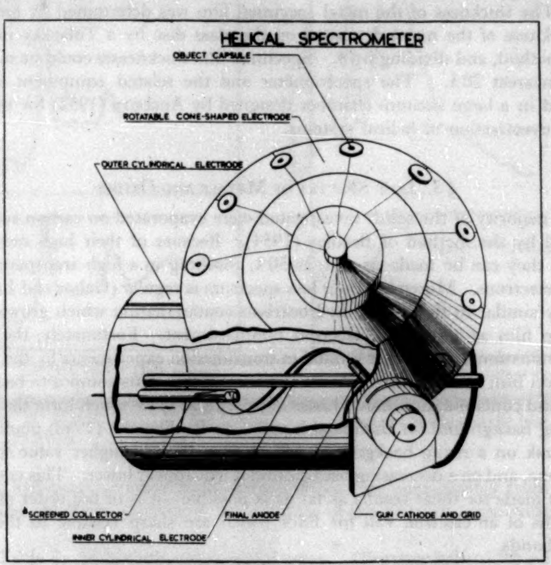

HELICAL SPECTROMETER
OBJECT CAPSULE
ROTATABLE CONE-SHAPED ELECTRODE
OUTER CYLINDRICAL ELECTRODE
SCREENED COLLECTOR
FINAL ANODE
GUN CATHODE AND GRID
INNER CYLINDRICAL ELECTRODE

Figure 1. Helical focusing velocity spectrometer.

A number of specimen films could be prepared and analysed without breaking the vacuum. This was convenient for rapid analysis of several specimens and, more important, necessary in order to reduce surface contamination of easily oxidized metals. Before a series of runs, 200-mesh electron microscope grids were covered with substrates, and mounted on a cone-shaped specimen electrode. Up to 12 grids were thus available for use at any time in the vacuum. A specimen film was prepared by evaporating the material vertically on to a grid, from evaporators positioned below the cone. The cone could then be rotated from outside the chamber, until the desired specimen film was accurately centred on the axis of the spectrometer electron gun, and the spectral analysis begun immediately. To complete the system, a rotatable glass disc was mounted halfway between the specimen cone and the evaporator, so arranged that the evaporated material deposited only on a small area of the disc and one specimen grid at a time. The thickness of the metal specimen film was determined by measuring the thickness of the metal deposited on the glass disc by a Tolansky interferometer method, and dividing by 4. Specimen film thicknesses could be measured to the nearest 20 Å. The spectrometer and the related equipment were all mounted in a large vacuum chamber designed by Ausburn (1952) for use in an earlier investigation of helical systems.

§ 3. Loss Spectra of Metals and Oxides

The majority of the solids investigated were evaporated on carbon substrates prepared by the method of Bradley (1954). Because of their high mechanical strength they can be made as thin as 50 Å, resulting in a high transparency for kilovolt electrons. Moreover, their loss spectrum is regular (Gabor and Jull 1955) and very similar to that of the hydrocarbon contamination which grows on the specimen film as a result of electron bombardment. Fortunately, the growth of contamination is somewhat limited in transmission experiments by the heating of the thin film by the electron beam (Ennos 1954). This composite backing of carbon and contamination shows losses of 5·5 ev and 23 ev which form the general scattering background for the metal losses. As Ruthemann (1948) pointed out, a loss peak on a rising background will show a slightly higher value than the correct one, and on a decreasing background it will appear lower. This correction has been made for these results as far as is possible—it is of the order of only a few tenths of an electron volt for lines which are sharp relative to the background bands.

Table 1 lists the losses observed in metals in the present investigation, which are divided into A and B type losses.

Table 1. Electron Energy Losses (ev)

Material	A-type					B-type					$\hbar\omega_p$
Al	14·6	29·2	43·8	58·4	7·30			7·0	20·5		15·9
Mg	10·0	20·3									11·0
Be	19·0	40·0	59·5								18·8
Si	17·1	34·0									16·9
C	23·0	46·0					5·0				23·5
Ag	23·8	44·8				3·3		6·8	14·3		9·0
Au	24·0	47·0						5·5	16·1	32·2	9·0

The A-type losses arise from the dominant loss process in each metal, while the B-type are less pronounced. The energy loss spectra for Al, Mg, Be and Si are illustrated in figures 2, 3, 4 and 5. There is general agreement between our A-type losses and those reported by other workers. Some discrepancies exist between our work and that of others for the B-type losses, with one additional loss reported in Be, Si (Marton and Leder 1954) and Mg (Kleinn 1954). For Ag and Au some authors do not find any of the losses below 20 ev, while others report only one (Marton, Leder and Mendlowitz 1955).†

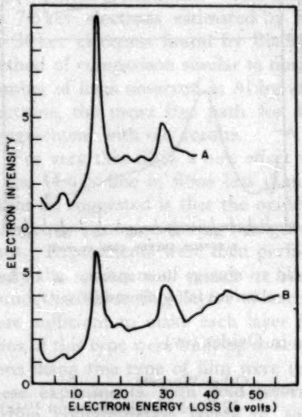


Figure 2. Energy loss spectra of 10 kev electrons in aluminium. A, 220 Å film. B, 370 Å film.

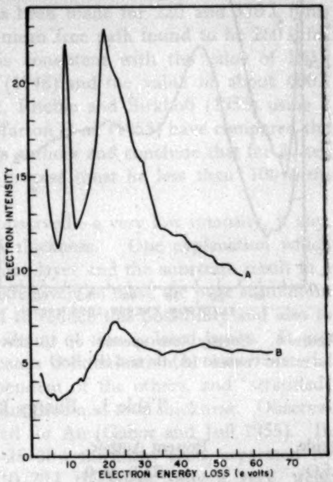


Figure 3. Energy loss spectra of 10 kev electrons in (A) Mg and (B) MgO.

The half-height widths of the A-type losses vary considerably from metal to metal and are shown in table 2, as multiples of the primary beam width.

Table 2. Half-widths of A-type Losses. (Primary half-width = 1.)

Metal	A loss (ev)	Half-width
Al	14·6	1·1
Mg	10·0	1·1
Be	19·0	6·2
Si	17·1	4·1
C	24·0	9·4
Ag	23·8	10·0
Au	24·0	9·4

Observations were made also on Al_2O_3, MgO, BeO and SiO_2. The first three were prepared by allowing the metal films to oxidize completely in air, and quartz films were prepared by direct evaporation. The observed losses are presented in table 3, along with the results of other authors. The spectra of MgO and BeO are shown in figures 3 and 4. For Al_2O_3 we find reasonable

---

† This review covers the recent experimental investigations of characteristic losses most thoroughly.

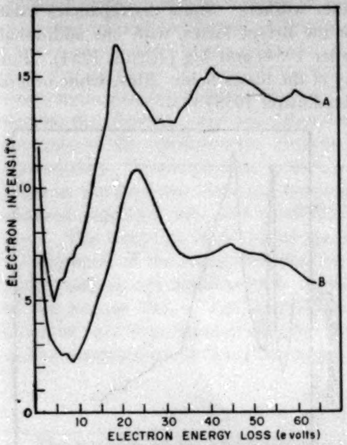


Figure 4. Energy loss spectra of 10 kev electrons in (A) Be and (B) BeO

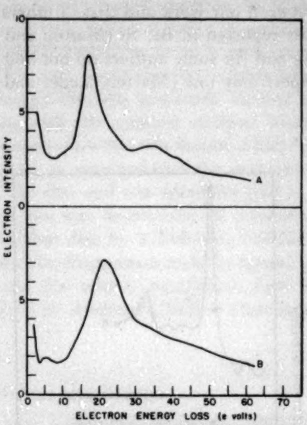


Figure 5. Energy loss spectra of 10 kev electrons in (A) Si and (B) SiO_2.

Table 3. Energy Losses in Oxides (ev)

Oxide	Present Author				Other Workers					
Al_2O_3			23·0	46·0				22·3	45·52	(Ruthemann 1948)
MgO	5·0	12·0	22·5		4·5	5·5	11·4	25·0		(Watanabe 1954)
BeO	6·5		23·0	45		5·7	16·5	28	57	(Watanabe 1954)
SiO_2	5·2		21·3			5·4	19·4			(Marton and Leder 1954)

agreement with other reported values but the MgO and BeO spectra are quite different than previously reported. The loss of 11·4 ev found by Watanabe in MgO was observed in several of our *partially* oxidized films of magnesium.

The spectra of Al and Mg do not bear any relation to their oxide spectra, while the losses observed for Be and Si are similar to the main losses observed for their oxides—the only difference being that the main oxide loss is shifted by 4 ev (Be) and 4·2 ev (Si) and is slightly broader. Marton and Leder (1954) first noted this similarity for Si and SiO_2 along with several other metals, quoting also the results of Watanabe (1954) for Mg and MgO. Our results for Mg and MgO do not confirm this similarity.

§ 4. THE VARIATION OF THE LOSS SPECTRUM WITH FILM THICKNESS

To determine the mean free path for the Al 14·6 ev line a study was made of the intensity of losses observed in films of different thicknesses. For single electron collisions, the probability of exciting a loss increases with the distance x

travelled by a primary electron in the solid. In a thickness dx, the probability of the electron losing one quantum is dx/λ where λ is the mean free path for the process. Then the probability $p_N(x)$ that one electron will lose N quanta in a thickness x can be shown to be

$$p_N(x) = \frac{1}{N!}\left(\frac{x}{\lambda}\right)^N e^{-x/\lambda}. \qquad \text{......(3)}$$

By comparing the area under a loss peak with the area under another for the same film, and knowing the film thickness, the mean free path can be obtained from equation (3). This comparison has been made for 220 and 370 Å films (figure 2) and the averaged value of the mean free path found to be 260 Å for 10 kev primary electrons. This result is consistent with the value of 180 Å for 7·6 kev electrons estimated by Lang (1948) and the value of about 600 Å for 30 kev electrons found by Blackstock, Ritchie and Birkhoff (1955) using a method of comparison similar to ours. Marton *et al.* (1955) have compared the number of lines observed in Al by various authors and conclude that for 30 kev electrons, the mean free path for this process must be less than 100 Å, in disagreement with our results.

For very thin films a new effect was observed—a very low intensity, if any, of the 14·6 ev line in films less than 30 Å thickness. One explanation which might be suggested is that the oxide surface layer and the substrate result in a sufficiently strong scattering background effectively to mask the pure aluminium lines. Experiments were then performed to reduce this possibility and also to study the spectrum of metals in films built up of thin isolated layers. It was found that separating layers of carbonaceous contamination or other material were sufficient to make each layer independent of the others, and 'stratified' films of this type were built up to several hundred a.u. total thickness. Observations using this type of film were reported for Au (Gabor and Jull 1955). In these experiments with gold layers of 25–30 Å thickness each, separated by carbon contamination layers of some 10–20 Å thickness, only a very weak loss line of 16·1 ev can be seen in films of as much as 170 Å aggregate gold thickness. This spectrum was compared with the spectrum of 110 Å gold film on a carbon backing which shows the 16·1 ev loss at least an order of magnitude stronger. (Since the prominent broad gold loss at 24·0 ev lies on top of the carbon maximum, it is difficult to detect any difference in this loss peak between a stratified and an unstratified film.)

The experiments with gold exclude the possibility of oxides affecting the spectrum, but the losses for gold which were weak in the stratified films are somewhat broader and less pronounced than the main losses in aluminium. Experiments with aluminium show that the strong 14·6 ev line is similarly weakened by building up the film of sufficiently thin layers. Moreover, all materials which were used as dividing layers appeared effective. The apparatus was provided with two evaporators and stratified films of the constituent materials could be built up by rapid alternate evaporation. Films of Al, with Au as a separating layer, were built up of various thicknesses and numbers of layers; figure 6 shows the spectrum of two such films. Curve A is that for 5 Al layers of 110 Å aggregate thickness separated by 4 thin Au layers. The only structure in the spectrum is a broad low maximum from 5 to 8 ev, a broad loss from 14·5 to 15 ev, as well as maxima at 23 ev and 45 ev. Curve B is the spectrum of two

Al layers of 120 Å total thickness separated by one 30 Å gold layer. Only a very weak 14·6 ev line appears. By way of comparison, figure 7 is the spectrum of a uniform Al film of 110 Å thickness, on to which have been evaporated

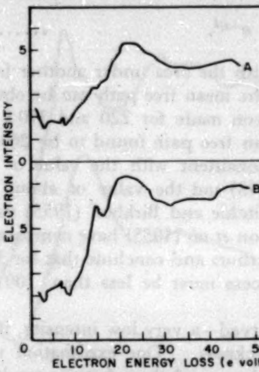


Figure 6. Energy loss spectra of 10 kev electrons in stratified aluminium films with gold as separating layers. A, 5 layers of aluminium of 110 Å aggregate thickness separated by 4 gold layers of about 40 Å total thickness. B, 2 layers of aluminium of 120 Å aggregate thickness separated by a 30 Å gold layer.

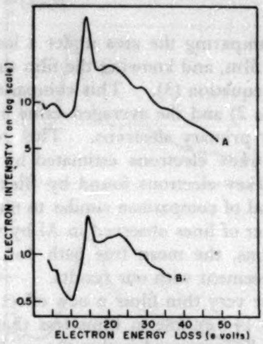


Figure 7. Energy loss spectra of 10 kev electrons in a continuous aluminium film with gold as a surface layer. A, 110 Å aluminium film. B, 110 Å aluminium film with gold surface layers of 200 Å aggregate thickness.

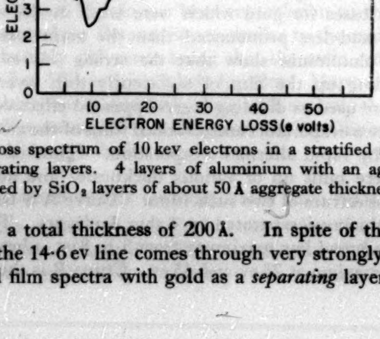


Figure 8. Energy loss spectrum of 10 kev electrons in a stratified aluminium film with SiO_2 as separating layers. 4 layers of aluminium with an aggregate thickness of 200 Å, separated by SiO_2 layers of about 50 Å aggregate thickness.

thin gold layers to a total thickness of 200 Å. In spite of the scattering in the gold *surface* layer the 14·6 ev line comes through very strongly, in sharp contrast to the Al stratified film spectra with gold as a *separating* layer. A similar effect resulted when SiO_2 was used as a separating layer as shown in figure 8. In all cases it appears that the interaction of the primary electrons with the 14·6 ev loss mechanism has been appreciably modified by dividing up the Al into thin layers of perhaps 50 Å and less, with a separating material.

§ 5. DISCUSSION

Our observed losses lend support to the suggestion of Pines (1955) that collective excitation could account for the A-type losses in Al, Mg, Be, Si and C. For a metal behaving as a free electron gas of density n, the collective loss $\hbar\omega_p$ is given by $3{\cdot}74\times10^{-11}\sqrt{n}$ ev; and the computed values (assuming all the valence electrons are free) are listed in table 1. As Pines has pointed out, the agreement is good between the calculated and observed losses in Al, Mg and Be —metals which would be expected to show free electron gas-like behaviour. Moreover, this agreement extends also to Si and C—solids with filled zones. This is possible for, as noted by Pines, the energy in a quantum of collective oscillation is greater than band separations for many solids, hence the usual distinctions between insulators and metals may not be generally applicable when considering collective excitation. Any collective losses in such metals as Ag and Au will be more difficult to identify tentatively, as interaction of the valence electrons with inner electrons can be expected to displace the free electron plasma loss.

Some of the observed properties of the losses show only that the observations are not irreconcilable with a collective interpretation. This is evident when comparing the spectrum of a metal with that of its oxide. The fact that the sharp lines observed in Al and Mg do not appear in their respective oxide spectra is as one would expect assuming free electron-like collective behaviour. On the other hand, while the observations on the similarity of the spectra of Be and BeO, and Si and SiO_2, may favour individual electron excitation, a collective interpretation cannot be ruled out for the losses in these solids. Clearly the conduction electrons cannot be responsible for losses in insulators but collective-like behaviour of core electrons might be expected in some cases as suggested by Gabor (1953, unpublished) and Pines (1953). A separation of losses of different origin does not appear possible by simply comparing the spectra of a metal and its oxide.

One possibility which might be considered in determining the origin of the losses is the observed value of the mean free path. Pines (1953) has determined the mean free path λ for collective excitation to be

$$\lambda=\frac{\hbar\omega_p m v^2}{4\pi n e^4 \ln(k_c v/\omega_p)} \qquad \ldots\ldots(4)$$

where v is the velocity of the incident electrons and n, as before, is taken as the number of valence electrons. k_c is the cut-off wave number for collective oscillations and has been given by Pines (1955) as

$$k_c \simeq 0{\cdot}35 k_0 \left(\frac{r_0}{a_0}\right)^{1/2} \qquad \ldots\ldots(5)$$

where k_0 is the wave number of an electron at the top of the Fermi level, a_0 is the Bohr radius, and r_0 is the interelectronic spacing, defined by

$$n=(\tfrac{4}{3}\pi r_0^{3})^{-1}. \qquad \ldots\ldots(6)$$

For the 14·6 ev line in Al, $k_c = 0{\cdot}89 \times 10^8\,\mathrm{cm}^{-1}$ and, with 10 kev electrons, $\lambda = 235$ Å. This is within 10% of the measured value, representing good agreement. The measurement of the mean free path cannot, however, be considered a sensitive test for the loss origin as pointed out by Marton *et al.* (1955) as both individual electron and collective excitation will result in mean free paths of the same order.

There still remain the observations on the initial non-linear rise in collision cross section for losses in thin stratified films of Au and Al. Clearly these could not be interpreted by single electron excitation, which one would expect to depend only on the aggregate metal thickness and not its structure. Nor can the results be accounted for by the Bohm–Pines collective excitation model. These authors have treated the excitation of plasma waves as an inelastic process in an extended medium, resulting in the same type of scattering cross section as would arise from individual electron excitation. The only model which appears appropriate to explain the results is that proposed by Gabor (1956). Gabor has recognized that for the interpretation of collective losses in thin films—as used in all experiments—representation of the incident electron beam as a plane monochromatic wave is not sufficient. For such films the interaction time τ depends only on the duration of the incident wave train, defined by the thermal velocity spread of the beam electrons, passing through the film. Moreover, for films in which the lateral grain size is greater than the geometrical coherent beam width (incoherent illumination) it was found that the interaction probability will be greater than the case for which the grains are coherently illuminated. As defined by Gabor the coherent beam width is given by half the electron wavelength λ_ε, divided by the angle subtended at the film by the image of the hot cathode. For 10 kev electrons and a collimation of the order of 10^{-3}–10^{-4} radians this gives coherence widths of 50–500 Å. Examination of electron micrographs and diffraction diagrams (Gabor and Jull 1955) shows that for the Au films used the grain size was greater than this range of coherent widths. Then for this case of incoherent illumination, Gabor has calculated the first-order interaction probability P as†

$$P = 5{\cdot}8 \times 10^{-4} \left(\frac{c}{v}\right) (\omega_p \tau) Q\left(\frac{a_\pi}{\frac{1}{2}\lambda_c},\ \frac{a}{a_\pi}\right) \qquad \ldots\ldots(7)$$

where a is the thickness of the film and $a_\pi = \pi v/\omega_p$ is a measure of film thickness in which an individual electron will 'see' only one half-cycle of collective oscillation ω_p. λ_c is the wavelength corresponding to the cut-off wave number k_c. $Q(a_\pi/\frac{1}{2}\lambda_c,\ a/a_\pi)$ illustrated in figure 9 (after Gabor), is determined by the summation of the contribution to the interaction probability of the individual excited modes. This Q-function does exhibit a marked departure from linearity —rising slowly at first, more rapidly at a thickness a_π, and finally reaching an asymptotic value after overshooting it slightly at $2a_\pi$. Neglecting 2nd and higher order transitions in very thin films, this Q-function alone determines the rise in interaction probability with increasing film thickness, since the interaction time is now independent of film thickness.

Gabor has made a comparison of the observations reported for Au, and the predicted non-linear increase in P. Since for these very thin films only first-order transitions are important, it is valid to compare the observations directly with the expected effect predicted by the Q-function. It was shown, however, that the observed non-linearity is too strong to be explained by the theory. The discrepancy, which lies outside experimental inaccuracies, can only be explained by assuming that the effective thickness of such thin layers is less than the nominal thickness, as regards collective oscillation. In conclusion Gabor has suggested that disturbed surface layers of several atomic layers would serve as the actual boundaries of collective waves in metal films; the results could then be readily explained.

---

† Correcting a numerical error in Gabor's eqn. 36. For coherent illumination the factor is $1{\cdot}34 \times 10^{-4}$.

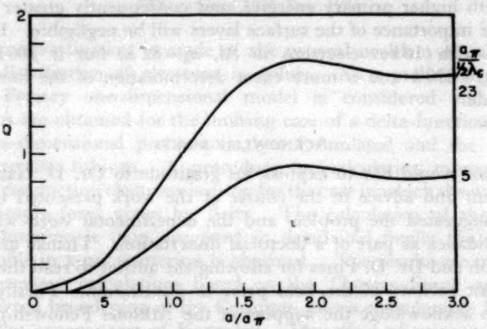


Figure 9. The collision probability function Q, as a function of the thickness and cut-off wavelength λ_c (after Gabor).

A better comparison with this theory would be expected for Al, for which the sharp loss tentatively identified with collective oscillation is the dominant inelastic process in the forward direction. Table 4 presents the ratio of the intensities observed for continuous films and films built up of thin layers. Also listed is the expected relative effect based on Gabor's theory. For the loss of 14·6 ev in Al and 10 kev electrons $a_\pi = 82$ Å. The cut-off wavelength predicted by Pines (1955) and confirmed by Watanabe (1956) is 7 a.u. The calculated relative intensities were taken for $a_\pi/\frac{1}{2}\lambda_c = 23$ from the Q-function (figure 9).

Table 4. Comparison of the Observed and Calculated Relative Intensities for Stratified and Continuous Aluminium Films

Observed				Calculated	
Stratified Film		Continuous Film	Relative	Relative	Corrected
No. of Layers	Total Thickness	Thickness (Å)	Intensity I_0 $\frac{I\text{ (continuous)}}{I\text{ (stratified)}}$	Effect I_c $\frac{I\text{ (continuous)}}{I\text{ (stratified)}}$	Relative Effect for Oxide
5	110	110	23	4·2	Very large
2	120	110	6	1·4	2
4	200	180	6	1·2	3

As for Au, there is still a stronger non-linearity than is predicted by the Q-function. For Al the presence of the oxidized surface layer must be considered; this would result in an actual thickness of perhaps 20 Å less than the measured thickness. For this reduced thickness the predicted effect is also shown in table 4—still not sufficient to account for the measured effect. If one assumes, as Gabor has done for Au, that disturbed boundary layers of several atomic layers are responsible for reducing the thickness for collective oscillations by another 15–20 Å, the results can be accounted for.

In conclusion, these experiments do undoubtedly show that the initial increase in P is non-linear, although the observations always show a stronger effect than can be accounted for by Gabor's theory. The theory predicts that the non-linearity will be observed up to thicknesses of the order of a_π'; this suggests operating with higher primary energies, and consequently greater values of a_π, for which the importance of the surface layers will be negligible. In the present experiments with 10 kev electrons in Al, $a_\pi = 82$ Å, but if 100 kev electrons were used $a_\pi = 230$ Å and a more exact determination of the non-linear effect could be made.

ACKNOWLEDGMENTS

The author would like to express his gratitude to Dr. D. Gabor for much encouragement and advice in the course of the work presented in this paper. Dr. Gabor suggested the problem and the experimental work was performed under his guidance as part of a doctorial dissertation. Thanks are also due to Dr. L. Marton and Dr. D. Pines for allowing the author to read the manuscripts of their recent reviews of the field prior to publication. Finally, the author would like to acknowledge the support of the Athlone Fellowship Committee, the Canadian National Research Council and the Defence Research Board during the period of this work..

REFERENCES

Ausburn, K. J., 1952, *M.Sc. Thesis*, University of London.
Blackstock, A. W., Ritchie, R. H., and Birkhoff, R. D., 1955, *Phys. Rev.*, **100**, 1078.
Bohm, D., and Pines, D., 1952, *Phys. Rev.*, **85**, 338.
Bradley, D. E., 1954, *Brit. J. Appl. Phys.*, **5**, 65.
Ennos, A. E., 1954, *Brit. J. Appl. Phys.*, **5**, 27.
Ferrell, R. A., 1956, *Phys. Rev.*, **101**, 554.
Gabor, D., 1951, *Proc. Phys. Soc.* B, **64**, 244; 1956, *Phil. Mag.*, (8) **1**, 1.
Gabor, D., and Jull, G. W., 1955, *Nature, Lond.*, **175**, 718.
Kleinn, W., 1954, *Optik, Stuttgart*, **11**, 226.
Lang, W., 1948, *Optik, Stuttgart*, **3**, 233.
Marton, L., and Leder, L. B., 1954, *Phys. Rev.*, **95**, 1345.
Marton, L., Leder, L. B., and Mendlowitz, H., 1955, *Advances in Electronics* (New York: Academic Press), **7**, 183.
Mollenstedt, G., 1949, *Optik, Stuttgart*, **5**, 499.
Pines, D., 1953, *Phys. Rev.*, **92**, 626; 1955, *Advances in Solid State Physics* (New York: Academic Press), Vol. 1.
Ruthemann, G., 1941, *Naturwissenschaften*, **29**, 648; 1948, *Ann. Phys., Lpz.*, **2**, 113.
Watanabe, H., 1954, *J. Phys. Soc. Japan*, **9**, 920; 1956, *Ibid.*, **11**, 112.
Wolff, P. A., 1953, *Phys. Rev.*, **92**, 18.

Momentum Distribution of Electrons in Solids

II : Some General Results with an Application to Metallic Lithium

By B. DONOVAN † and N. H. MARCH ‡

† Department of Physics, Bedford College, London

‡ Department of Physics, The University, Sheffield

MS. received 12*th June* 1956

Abstract. An investigation is made of the general problem of calculating the momentum distribution of electrons in solids. As a preliminary illustration the Kronig–Penney one-dimensional model is considered and momentum eigenfunctions are obtained for the limiting case of a delta-function potential.

The three-dimensional problem is then formulated and the methods are applied to metallic lithium. A procedure for calculating momentum eigenfunctions for conduction electrons is given for the case in which the wave functions are of the simple form given by Seitz. The calculation of the momentum distribution function is then discussed in detail and finally the shape of the Compton profile in x-ray scattering is obtained. The results are in good agreement with previous calculations based on the Thomas–Fermi approximation but the predicted broadening of the Compton line is considerably less than that obtained in the experiments of Kappeler. The effect of using a more refined wave function is investigated but it appears unlikely that the discrepancy can be resolved in this way.

§ 1. Introduction

In part I of this series (March 1954) attention was drawn to experimental results giving the shape of the Compton profile in x-ray scattering from metallic lithium and beryllium. The very marked broadening found by Kappeler (1936) and Dumond (1933) respectively for these two metals presents an interesting theoretical problem which ought to yield information about the accuracy and general validity of present theoretical treatments of metals. It should be understood here that the broadening referred to above is that found by comparing the experimental half-widths for these metals with the theoretical half-widths for the corresponding isolated atoms which can be found from a paper by Duncanson and Coulson (1945). The half-width of the Compton profile is increased by some 350% according to Kappeler for Li whilst for Be the increase is about 160% from the experimental results of Dumond.

In I the theoretical results were obtained by using the Thomas–Fermi approximation to calculate the momentum distribution and whilst it was possible in this way to demonstrate conclusively that one would expect a marked broadening of the Compton line by binding atoms in a crystal the results left much to be desired from a quantitative point of view. Surprisingly enough, this was particularly so for Li, in which case a broadening of 100% was anticipated, much less than Kappeler found experimentally.

In this paper we take up the general problem of calculating the momentum distribution of electrons in solids. We shall first of all deal with a one-dimensional case, using the Kronig–Penney model to illustrate the methods we shall employ. This is considered in § 2 where we also discuss the relation of our work to that of Slater, who alone seems to have given any attention to the sort of problem we meet here. Then in § 3 the general three-dimensional problem is considered and a practical application of the results is made for Li. In § 4 we discuss the method of finding the momentum distribution function which is needed in the calculation of the shape of the Compton profile. The calculations are carried out explicitly for Li and the Compton profile obtained. In § 5 we discuss the connection between the present results and those obtained previously using the Thomas–Fermi approximation and make a comparison with experiment, while in § 6 we briefly consider the effect of refining the wave functions we have used.

§ 2. MOMENTUM EIGENFUNCTIONS IN ONE-DIMENSIONAL PROBLEMS

We begin by considering a general one-dimensional periodic potential. We are interested in solutions of the Schrödinger equation of the Bloch form

$$\psi_k(x) = u_k(x) \exp(ikx) \qquad \ldots\ldots(1)$$

where the function $u_k(x)$ is periodic with the period of the potential. Since u_k is periodic we can expand it in a Fourier series and write

$$u_k = \sum_{n=-\infty}^{\infty} v_{nk} \exp\left(\frac{2\pi i n x}{a}\right) \qquad \ldots\ldots(2)$$

and thus we have

$$\psi_k = \sum_{n=-\infty}^{\infty} v_{nk} \exp\left[i\left(k + \frac{2\pi n}{a}\right) x\right] \qquad \ldots\ldots(3)$$

where a is the fundamental periodicity, that is

$$u_k(x+a) = u_k(x).$$

The physical interpretation of equation (3) is now clear. Each term in the expansion of ψ_k is an eigenfunction of the momentum with eigenvalue $p = k + 2\pi n/a$.† In other words, in equation (3) we have the expansion of the space wave function in terms of eigenfunctions of momentum and by the usual quantum mechanical interpretation it follows that when the wave function defined by equation (3) is suitably normalized the coefficients v_{nk} are such that $|v_{nk}|^2$ gives the probability that an electron with wave number k shall have a momentum $k + 2\pi n/a$. (For a completely free electron, for example, v_{0k} will be equal to unity and all other coefficients will be identically zero.) The coefficients v_{nk} thus define the momentum eigenfunction for the particular electron described by the wave number k and show us that there is a finite probability of finding any one of an infinite number of discrete values for the momentum. Although the idea of associating a momentum eigenfunction with any particular electron is familiar, by analogy with atomic and molecular theory, there is another way of looking at the problem for solids which, as Slater has emphasized, is particularly convenient. Here we have the case when a band is described in terms of a continuous range of wave numbers (or wave vectors in the three-dimensional case) and whilst for a particular value of k the momentum eigenfunction is defined only at discrete points in momentum space, if we now consider all the values of k within a band we can regard the separate momentum eigenfunctions as building up a continuous function which gives us complete information about all the wave functions within the band and which can conveniently be referred to as the momentum eigenfunction of the band.

---

† Atomic units are used throughout this paper.

The problem of finding the coefficients v_{nk} which constitute the momentum eigenfunction can be approached in either of two ways. The first, and usually the more straightforward, is to start from the space wave function $\psi_k(x)$ and transform it into momentum space, whilst the second is to set up appropriate equations for the momentum eigenfunctions and to solve them directly. In this paper we shall be concerned solely with the first method.

Thus we assume $\psi_k(x)$ or $u_k(x)$ to be known. If we multiply equation (2) by $\exp(-2\pi imx/a)$ and integrate over a fundamental period we obtain

$$v_{mk}=\frac{1}{a}\int_0^a u_k(x)\exp\left(\frac{-2\pi imx}{a}\right)dx. \qquad \ldots\ldots(4)$$

Use of Kronig–Penney Delta-Function Potential

We shall illustrate the methods used in this paper by considering the one-dimensional periodic potential first proposed by Kronig and Penney (1931). Even this potential does not lead to simple wave functions in general and therefore in the application which follows we shall deal only with the case when the potential hills become infinitely high and infinitely thin, that is we deal with a periodic sequence of delta-functions. The consequences of such a model have been discussed in great detail by Hutner and Saxon (1949).

In this case we may write

$$u_k(x)=C\exp[i(-k+\beta)x]+D\exp[i(-k-\beta)x] \qquad \ldots\ldots(5)$$

where $\beta^2=2E$, E being the energy corresponding to the wave number k, and

$$D=-C\frac{1-\exp[-i(k-\beta)a]}{1-\exp[-i(k+\beta)a]}.$$

Further, if we require that

$$\lim_{L\to\infty}\left[\int_0^L \psi_k{}^*\psi_k dx \Big/ \int_0^L dx\right]=1$$

which then ensures (Slater 1934) that

$$\sum_m |v_{mk}|^2=1$$

we find

$$CC^*=\frac{1}{2}\left[\frac{1-\cos[(k+\beta)a]}{1-\cos ka\cos\beta a+M^{-1}(\cos ka-\cos\beta a)^2}\right]. \qquad \ldots\ldots(6)$$

Here, M is related to β by the equation

$$(M\sin B/B)+\cos B=\cos P$$

where we have put $\beta a=B$, $pa=ka+2\pi m=P$.† It can then be shown after somewhat lengthy calculation starting from equation (4) that

$$|v_{mk}|^2=\frac{4M^2\sin B}{[\sin B(1+M/B^2)-(M/B)\cos[B]B^2-P^2]^2}. \qquad \ldots\ldots(7)$$

Using this expression we can plot out the momentum eigenfunctions for all the energy bands. In figure 1 we illustrate the result represented in equation (7) by plotting $|v_{mk}|^2$ for various special cases. Curves 1–3 show $|v_{mk}|^2$ for the first three bands assuming (following Kronig and Penney) that $M=3\pi/2$. It is of interest to compare these curves with those obtained by Slater (1952, fig. 4) in his discussion of the one-dimensional Mathieu problem; some similarity is evident. The general effect of varying M can be seen fairly readily. If we make M large, then as Kronig and Penney pointed out, the regions of allowed energies contract eventually into points defined by $B=\pi$, 2π, 3π, etc. We show in curves 4–6 the results for the first three 'bands' in the limit $M\rightarrow\infty$. It is easy to verify that these are just the momentum eigenfunctions for the first three levels in a square well potential. The opposite limit $M\rightarrow 0$ leads to the free electron case. We illustrate the effect of diminishing M by plotting the momentum eigenfunction for the first band, taking $M=3\pi/20$, in curve 7. Already we have a close approximation to a free electron case, with $|v_{mk}|^2$ practically constant and very near to unity and then suddenly dropping to zero. The results for higher bands in all cases are quite illuminating but in the interests of brevity will not be discussed further.

---

† Our notation differs slightly from that of Kronig and Penney. In particular our M corresponds to their P.

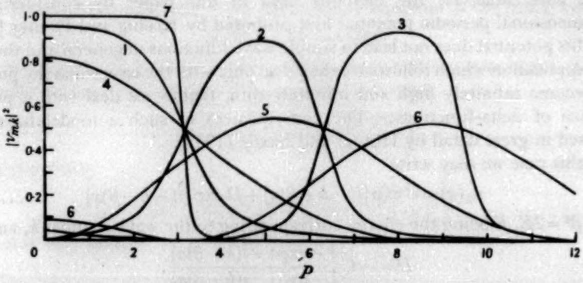


Figure 1. Momentum eigenfunctions for Kronig–Penney delta-function potential. Curves 1–3 : first three 'bands' for $M=3\pi/2$; curves 4–6 : first three 'bands' for $M=\infty$; curve 7 : first band for $M=3\pi/20$.

§ 3. FORMULATION FOR GENERAL THREE-DIMENSIONAL CASE

In order to calculate the momentum eigenfunctions we shall suppose that we know the space wave function $\psi_{\mathbf{k}}(\mathbf{r})$ describing a valence electron in a state defined by the wave vector $\mathbf{k}$. According to Bloch's theorem we may write

$$\psi_{\mathbf{k}}(\mathbf{r})=u_{\mathbf{k}}(\mathbf{r})\exp(i\mathbf{k}\,.\,\mathbf{r}) \qquad \ldots\ldots(8)$$

where $u_{\mathbf{k}}(\mathbf{r})$ is periodic with the period of the lattice. This may, therefore, be expanded in a triple Fourier series, making use of the reciprocal lattice. If $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$ are the basic vectors in the reciprocal lattice and $\mathbf{G}$ is a vector defined by

$$\mathbf{G}=2\pi(g_1\mathbf{b}_1+g_2\mathbf{b}_2+g_3\mathbf{b}_3) \qquad \ldots\ldots(9)$$

where g_1, g_2, g_3 are integers, then we have

$$u_{\mathbf{k}}(\mathbf{r})=\sum_{g_1g_2g_3=-\infty}^{\infty} v_{\mathbf{Gk}}\exp(i\mathbf{G}\,.\,\mathbf{r}). \qquad \ldots\ldots(10)$$

It follows that

$$\psi_{\mathbf{k}}(\mathbf{r})=\sum_{g_1g_2g_3=-\infty}^{\infty} v_{\mathbf{Gk}}\exp[i(\mathbf{k}+\mathbf{G})\,.\,\mathbf{r}] \qquad \ldots\ldots(11)$$

which is again an expansion in eigenfunctions of momentum. As in the one-dimensional case it therefore follows that $v_{\mathbf{Gk}}$ is the momentum eigenfunction for the electron in state $\mathbf{k}$, or as before by allowing $\mathbf{k}$ to run over all values within a band we can consider the continuous function which is thereby formed as the momentum eigenfunction of the band.

For any particular vector, $\mathbf{G}'$ say, we find $v_{\mathbf{Gk}}$ by multiplying both sides of equation (10) by $\exp(-i\mathbf{G}'.\mathbf{r})$ and integrating over a unit cell of volume V. In this way we obtain

$$v_{\mathbf{G'k}}=\frac{1}{V}\int_{\text{unit cell}} u_{\mathbf{k}}(\mathbf{r})\exp(-i\mathbf{G}'.\mathbf{r})d\tau. \qquad \ldots\ldots(12)$$

For this to give the momentum eigenfunction directly, the space wave function must be normalized so that

$$\int_{\text{unit cell}} \psi_{\mathbf{k}}^{*}\psi_{\mathbf{k}}d\tau=V. \qquad \ldots\ldots(13)$$

3.1. *Application to Seitz Wave Function*

We now apply equation (12) to Li, where we shall suppose it is sufficiently accurate to use the simple wave function due to Seitz (1935). We shall look more closely into this approximation later. This wave function can be written in the form

$$\psi_{\mathbf{k}}(\mathbf{r})=u_0(r)\exp(i\mathbf{k}.\mathbf{r}). \qquad \ldots\ldots(14)$$

Since the function $u_0(r)$ is determined using the sphere approximation we shall expect to take the integration in equation (12) not over the cellular polyhedron but over the sphere of equal volume. We can now use Bauer's expansion for a plane wave in an infinite series of spherical waves, namely

$$\exp(-i\mathbf{G}.\mathbf{r})=\sum_{j=0}^{\infty}(2j+1)(-i)^{j}\mathrm{P}_j(\cos\theta)f_j(Gr) \qquad \ldots\ldots(15)$$

where θ is the angle between $\mathbf{r}$ and $\mathbf{G}$ and the functions $f_j(Gr)$ are defined by

$$f_j(Gr)=\left(\frac{\pi}{2Gr}\right)^{1/2}\mathrm{J}_{j+1/2}(Gr),$$

J denoting a Bessel function in the usual notation. Performing the integration over angles we find

$$v_{\mathbf{Gk}}=\frac{1}{V}\int_0^{r_s}u_0(r)f_0(Gr)4\pi r^2dr \qquad \ldots\ldots(16)$$

where $f_0(Gr)=\sin Gr/Gr$, and r_s is the radius of the atomic sphere. This is the formula we have used to calculate the momentum eigenfunctions for Li. It is clear that the integral is not in fact dependent on $\mathbf{k}$ because of the special type of wave function used and this, of course, leads to considerable simplification. Having calculated the momentum eigenfunction for a particular electron in the band we then know it for all other electrons.

3.2. *Evaluation of Momentum Eigenfunctions*

In order to obtain the momentum eigenfunctions we need to enumerate the vectors in the reciprocal lattice. For a body-centred cubic lattice we have $|\mathbf{b}_1|=|\mathbf{b}_2|=|\mathbf{b}_3|$ and hence

$$\mathbf{G}=0{\cdot}957(g_1\mathbf{i}_1+g_2\mathbf{i}_2+g_3\mathbf{i}_3) \qquad \ldots\ldots(17)$$

(see, for example, Parmenter 1952), where $\mathbf{i}_1$, $\mathbf{i}_2$ and $\mathbf{i}_3$ are mutually orthogonal unit vectors and $g_1+g_2+g_3$ is an even integer. The formula (16) clearly involves only the lengths of these vectors, which are shown in the table. The evaluation of the integral can then be carried out and in this way the results shown in the table have been obtained.

Momentum Eigenfunction for Lithium

G	n_G	v_{Gk}	$\|v_{Gk}\|^2$
0	1	−0·947	0·897
1·353	12	+0·0442	0·00195
1·914	6	+0·124	0·0154
2·344	24	+0·0366	0·00134
2·707	12	−0·0136	0·000185
3·026	24	−0·0130	0·000169
3·315	8	+0·00786	0·0000618
3·581	48	+0·0247	0·000609
3·828	6	+0·0285	0·000811
4·060	24	+0·0214	0·000457
4·280	24	+0·00929	0·0000863
4·489	24	−0·00156	0·00000244

A number of comments are worth making in connection with the table. First of all, from a computational point of view, we found it useful to obtain v_{Gk} for the longer reciprocal lattice vectors by fitting a relatively simple analytical form to the Seitz wave function. This procedure is desirable from a practical point of view as the integrand oscillates rapidly, making the usual methods of numerical integration laborious to apply. The analytical form adopted is

$$\left.\begin{array}{ll} u_0(r)=1{\cdot}899[3{\cdot}547\exp(-2{\cdot}26r)-r\exp(-0{\cdot}658r)], & 0<r\leqslant 1{\cdot}80 \\ ru_0(r)=0{\cdot}1446r^2-1{\cdot}8636r+1{\cdot}1670 & ,\ 1{\cdot}80<r\leqslant r_s \end{array}\right\}\quad\ldots\ldots(18)$$

and the first expression is shown in figure 2 together with the Seitz wave function. It is doubtless possible to obtain a somewhat better fit by readjusting parameters but this is adequate for our purposes. For $r>1{\cdot}80$, the analytical form and the Seitz function are indistinguishable to graphical accuracy. Secondly, we see immediately from the table why a straightforward expansion of the wave function in plane waves is not very useful from a practical point of view, as has often been noted. This is because of the slow convergence of the v_{Gk}'s—although v_{0k},

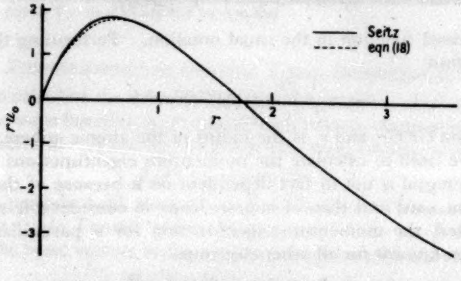


Figure 2. Comparison of analytical wave function with Seitz function.

Note. Author's correction: the figures on each of these scales should read 0, 0·5, 1 . . . . . instead of 0, 1, 2 . . . . .

is very much larger than any other v_{Gk}, nevertheless, many of the higher values are of the same order of magnitude. It should also be remembered that there is a number n_G of reciprocal lattice vectors with the same length G and we show these numbers also in the table.

The values of $|v_{Gk}|^2$ are also illuminating. We see that the probability of finding an electron of wave vector **k** with momentum **k** is 0·897, very much larger of course than the probability of finding any higher value. If we began with a suitably normalized wave function with the correct symmetry of the lattice then we should have the relation

$$\sum_G |v_{Gk}|^2 = 1, \qquad \ldots\ldots(19)$$

simply expressing the fact that the probability of finding the electron with one of its infinite number of possible discrete values of the momentum is unity. If we form this sum numerically from the results given in the table we find

$$\sum |v_{Gk}|^2 \sim 1{\cdot}1.$$

The deviation from unity appears, therefore, to have arisen chiefly because of the sphere approximation employed, although the possibility of small errors arising from the fitted functions (18) cannot be entirely excluded. We do not believe that our essential results will be in any way seriously affected by these approximations.

Finally, whilst dealing with the momentum eigenfunctions, it is interesting to examine the way in which the detailed form of the Seitz wave function is built up from the expansion in plane waves. To do this, we could go back to equation (10), insert the appropriate v_{Gk} values and see how $u_{\mathbf{k}}(\mathbf{r})$ is obtained. However, we calculated the v_{Gk} results from a spherically symmetrical wave function and

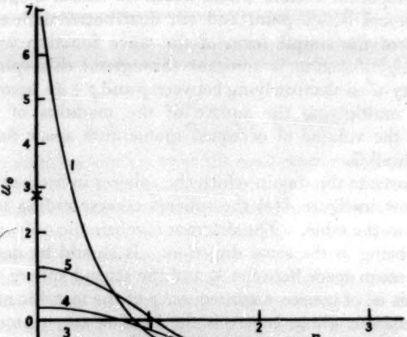


Figure 3. Illustration of approximation to Seitz function obtained from plane wave expansion. Curve 1 : Seitz wave function ; curve 2 : 1 plane wave ($G=0$); curve 3: 13 plane waves ($G=0$, 1·353); curve 4 : 19 plane waves ($G=0$, 1·353, 1·914); curve 5 : 43 plane waves ($G=0$, 1·353, 1·914, 2·344). ×: value at origin with 213 plane waves.

thus it seems more appropriate to average $u_{\mathbf{k}}(\mathbf{r})$ in equation (10) over angles. This amounts to replacing the plane wave $\exp(i\mathbf{G}.\mathbf{r})$ by $f_0(Gr)$. Thus we see that the first term is simply

$$v_{0\mathbf{k}}f_0(0)=v_{0\mathbf{k}}=-0{\cdot}947.$$

We show in figure 3 the way in which we approach the Seitz function as we add further terms. Curve 5 shows the result when we include all reciprocal lattice vectors up to $G=2{\cdot}344$, corresponding to the inclusion of 43 plane waves. It can be seen that we have a totally inadequate approximation to the Seitz function which is shown for comparison. If we add in all the plane waves corresponding to values of G recorded in the table, 213 in all, the value of u_0 at the origin (indicated by a cross in figure 3) is still less than half of the Seitz value. These results demonstrate very forcibly the need for modifying the method of expanding in plane waves and for the introduction by Herring (1940) of the orthogonalized plane wave method.

§ 4. CALCULATION OF MOMENTUM DISTRIBUTION FUNCTION

The momentum eigenfunctions are not themselves sufficient to enable us to calculate the required shape of the Compton line. We must now proceed to find the momentum distribution function $I(p)$ which is such that $I(p)dp$ gives the probability of finding an electron with momentum of magnitude between p and $p+dp$.

First of all, we have the result that the number of states per unit volume of $\mathbf{k}$ space per unit volume of metal is $1/8\pi^3$. Also, for every vector $\mathbf{k}$ lying inside the Fermi surface, which we take to be a sphere of radius k_0 $(=0{\cdot}589)$, we obtain an allowed value of the momentum by forming the sum of $\mathbf{k}$ and any vector $\mathbf{G}$, as can be seen from equation (11). For a given vector $\mathbf{G}$ in the reciprocal lattice, we have thus momentum vectors whose heads lie within a sphere of radius k_0 around the reciprocal lattice point and are distributed with uniform density. Further, because of the simple form of the wave function we have adopted, the momentum eigenfunction is constant throughout this sphere. Hence we find the probability of an electron lying between p and $p+dp$, apart from a normalizing factor, by multiplying the square of the modulus of the momentum eigenfunction by the volume of occupied momentum space for each vector $\mathbf{G}$ and summing over all $\mathbf{G}$.

In order to illustrate the way in which the spheres in momentum space begin to overlap we show in figure 4 (*a*) the spheres corresponding to the first seven values of G given in the table. The diagram is schematic only, as we have taken the vectors $\mathbf{G}$ as being in the same direction. It should be noted that there is a region of momentum space between k_0 and the second sphere which is entirely unoccupied. This is, of course, a consequence of the metallic nature of lithium. For a particular sphere whose centre is the head of the vector $\mathbf{G}$, the volume cut off between spheres of radii p and $p+dp$ (see figure 4 (*b*)) is

$$\pi\left[\left(\frac{k_0^2}{G}-G\right)p+2p^2-\frac{p^3}{G}\right]dp. \qquad \ldots\ldots(20)$$

We can now write down immediately the momentum distribution function $I(p)$ for the range $0<p\leqslant k_0$. The result is

$$I(p)dp=\frac{|v_{0\mathbf{k}}|^2\,4\pi p^2dp}{4\pi k_0^3/3}$$

or

$$I(p)=13{\cdot}2p^2$$

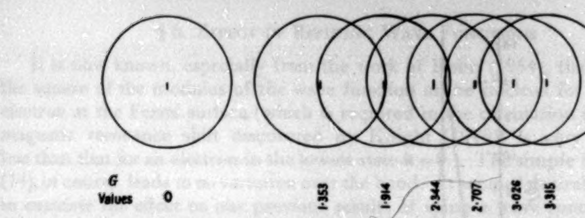


Figure 4 (*a*). Diagram (to scale) illustrating system of overlapping spheres in momentum space.

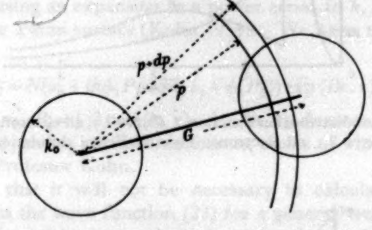


Figure 4 (*b*). Diagram showing notation used in equation (20) for volume of occupied momentum space.

whereas the free electron result is simply

$$I(p) = 3p^2/k_0{}^3 = 14{\cdot}7p^2.$$

Continuing in this way we build up the function $I(p)$ for higher values of the momentum p.

It should be noted that, so far, we have only considered the conduction electrons. We must now take account of the contribution from the 1s electrons. This can be conveniently obtained from the work of Duncanson and Coulson (1945) and may be written

$$I_{1s}(p) = \frac{32\alpha^5 p^2}{\pi(p^2+\alpha^2)^4}\,; \quad \alpha = 2{\cdot}69. \qquad \ldots\ldots(21)$$

From these two distributions we form the total momentum distribution function by taking two-thirds of $I_{1s}(p)$ and adding it to one-third of the momentum distribution function for the conduction electrons, and the final result is shown in curve 1 of figure 5. For comparison the result obtained in I by means of the Thomas–Fermi approximation is shown in curve 2.

Knowing the momentum distribution function we can now compute the Compton profile. This can be conveniently written in the following form applicable for all scattering angles and all incident wavelengths (Duncanson and Coulson 1945):

$$J(q) = \frac{1}{2}\int_q^\infty \frac{I(p)}{p}\,dp. \qquad \ldots\ldots(22)$$

The results obtained for $J(q)$ by numerical integration are shown in curve 1 of figure 6, whilst curve 2 represents the results of the Thomas–Fermi approximation. The half-width of the profile is a convenient quantity to use for comparison with experiment and the value is found to be 1·0 atomic unit.

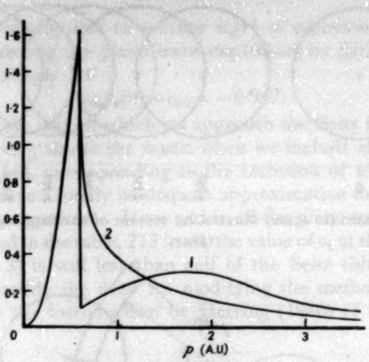


Figure 5. Momentum distribution function. Curve 1 : all electrons, present calculation; curve 2 : all electrons, Thomas–Fermi calculation.

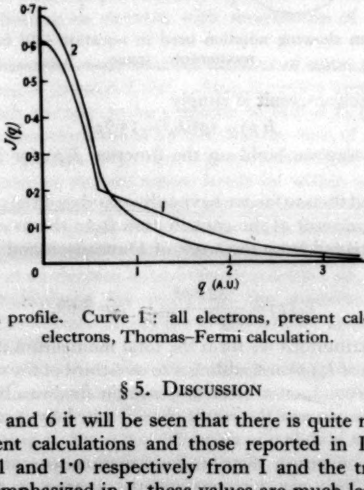


Figure 6. Compton profile. Curve 1 : all electrons, present calculation; curve 2 : all electrons, Thomas–Fermi calculation.

§ 5. DISCUSSION

From figures 5 and 6 it will be seen that there is quite reasonable agreement between the present calculations and those reported in I. In particular the half-widths are 1·1 and 1·0 respectively from I and the treatment given here. However, as was emphasized in I, these values are much less than that given by Kappeler (2·6). It is clear from our work that in order to obtain a result in agreement with Kappeler's the value of $I(p)$ at the peak, corresponding to $p=k_0$, must be very markedly reduced, and thus a bigger contribution to $I(p)$ must come at larger momenta. From our work it is clear that the magnitude of $I(p)$ at $p=k_0$ is determined by the momentum eigenfunction of an electron at the Fermi surface, and it is certainly here that we might expect the Seitz wave function to be least accurate. For this reason we shall consider in § 6 the use of a refined wave function due to Kohn.

§ 6. Effect of Refining Wave Functions

It is now known, especially from the work of Kohn (1954), that the value of the square of the modulus of the wave function at the nucleus for a conduction electron at the Fermi surface (which is required in the calculation of the nuclear magnetic resonance shift discovered by Knight (1949)) is very considerably less than that for an electron in the lowest state $\mathbf{k}=0$. The simple Seitz function (14), of course, leads to no variation over the band. It seemed desirable, therefore, to examine the effect on our previous results of using a wave function which is more accurate than the Seitz function at the Fermi surface.

As was first proposed by Bardeen (1938), it is possible to improve the Seitz wave function by using an expansion in a power series in k, and this appears to be still useful at the Fermi surface (Kohn 1954). We have, therefore, examined the wave function

$$\psi_{\mathbf{k}}(\mathbf{r})=N(u_0+ik\phi_1P_1+k^2\{\phi_0+\phi_2P_2\})\exp(i\mathbf{k}\,.\,\mathbf{r}) \qquad \ldots\ldots(23)$$

where $P_1=\cos(\mathbf{k}\,.\,\mathbf{r})$, $P_2=\frac{1}{2}\{3\cos^2(\mathbf{k},\mathbf{r})-1\}$, $(\mathbf{k},\mathbf{r})$ denotes the angle between $\mathbf{k}$ and $\mathbf{r}$ and u_0, ϕ_0, ϕ_1 and ϕ_2 are radial functions, tables of which have been made available to us by Professor Kohn.

We shall find that it will not be necessary to calculate the momentum eigenfunctions from the wave function (23) for a general vector $\mathbf{G}$ although this in fact is possible, if very laborious. As we have already noted, the most important region is the central sphere of momentum space, corresponding to $\mathbf{G}=0$, and particularly the neighbourhood of the Fermi surface. Making the sphere approximation as usual we find, by making use of equations (12), (15) and (23) and the orthogonality properties of the Legendre polynomials,

$$v_{0\mathbf{k}}=\frac{N}{V}\int(u_0+k^2\phi_0)d\tau. \qquad \ldots\ldots(24)$$

Evaluating the various integrals numerically we find that $v_{0\mathbf{k}}$ varies insignificantly over the occupied part of the band. Thus, at least on the basis of this wave function, the result obtained from the simple Seitz wave function seems well substantiated. Undoubtedly, if we were now to repeat the calculations for other values of $\mathbf{G}$ we should find some variation of $v_{\mathbf{G}\mathbf{k}}$ with $\mathbf{k}$ in general. But it seems clear from our work that such modifications will make only minor changes in the final result for the Compton profile, and we have not proceeded further.

§ 7. Conclusion

From the present work it would therefore appear that a real discrepancy exists between theory and experiment. We cannot, of course, be completely certain of the accuracy of our results at the Fermi surface, which is a very important region, but calculations using the wave function (23) entirely support the conclusions reached on the basis of the simpler function (14) as well as the results obtained from the Thomas–Fermi calculations of I. The possibility that changes in the shape of the Fermi surface can markedly alter the results should not be entirely discounted, although we do not believe this to be likely. What does seem clear from the present investigation is that further experimental work along the lines of that carried out by Dumond (1933) and Kappeler (1936) would be of very considerable interest.

ACKNOWLEDGMENTS

We are greatly indebted to Professor F. Seitz for sending us a table of the wave function (14) and to Professor W. Kohn for making the improved wave function (23) available to us.

REFERENCES

BARDEEN, J., 1938, *J. Chem. Phys.*, **6,** 367.
DUMOND, J. W. M., 1933, *Rev. Mod. Phys.*, **5,** 1.
DUNCANSON, W. E., and COULSON, C. A., 1945, *Proc. Phys. Soc.*, **57,** 190.
HERRING, C., 1940, *Phys. Rev.*, **57,** 1169.
HUTNER, R. A., and SAXON, D. S., 1949, *Philips Res. Rep.*, **4,** 81.
KAPPELER, H., 1936, *Ann. Phys., Lpz.*, **27,** 129.
KNIGHT, W. D., 1949, *Phys. Rev.*, **76,** 1259.
KOHN, W., 1954, *Phys. Rev.*, **96,** 590.
KRONIG, R. DE L., and PENNEY, W. G., 1931, *Proc. Roy. Soc.* A, **130,** 499.
MARCH, N. H., 1954, *Proc. Phys. Soc.* A, **67,** 9.
PARMENTER, R. H., 1952, *Phys. Rev.*, **86,** 552.
SEITZ, F., 1935, *Phys. Rev.*, **47,** 400.
SLATER, J. C., 1934, *Rev. Mod. Phys.*, **6,** 209 ; 1952, *Phys. Rev.*, **87,** 807.

Ageing of the Dielectric Properties of Barium Titanate Ceramics

By K. W. PLESSNER

British Dielectric Research Ltd., 38 Wood Lane, London, W.12

Communicated by J. B. Birks; MS. received 7th May 1956 *and in revised form* 18*th June* 1956

Abstract. The permittivity and power factor of ceramic barium titanate and barium titanate solid solutions have been measured as a function of time, starting from the time of cooling the material through the Curie point. Both parameters decrease linearly with $\log t$ over periods extending to several years. A theory is proposed which postulates that a large contribution is made to the permittivity by the motion of domain walls, and that the domain walls are gradually moved to more stable and less polarizable states by thermal agitation. A wide distribution of activation energies, reflecting the random character of obstacles to domain wall motion in a ceramic, yields the $\log t$ law. Domain walls moving over low-energy humps are considered responsible for the power factor decay.

§ 1. INTRODUCTION

The gradual diminution with time of the permittivity of barium titanate type ceramics has been known for some years. This 'ageing' effect starts when the ceramic is cooled through the Curie point, proceeds rapidly at first and continues at a gradually diminishing rate for a period of months or years. The ageing cycle can be repeated by heating the ceramic above its Curie point for a short time. The effect is, therefore, connected with the ferroelectric nature of $BaTiO_3$, and a number of authors have attributed ageing to the gradual rearrangement of domains into more stable configurations (Bogoroditskiĭ and Verbitskaya 1952, Novosiltsev *et al.* 1952). Mason (1955) has given a detailed theory of the ageing of the piezoelectric properties of Ba–Pb and Ba–Pb–Ca titanates, which will be discussed below.

The present paper describes measurements of the ageing of $BaTiO_3$ ceramics of reasonable purity. The results show that the permittivity decreases linearly with $\log t$, where t is the time, over several decades.† Such a $\log t$ relation had been suggested earlier by Marks (1948). A theory is developed in this paper, in which a $\log t$ relationship is derived from a very wide distribution of the activation energies governing the motion of domain walls. An analogous mechanism has been proposed by Street and Woolley (1949) to account for the phenomenon of magnetic viscosity in permanent magnet materials.

§ 2. EXPERIMENTAL

The specimens used were in the form of small discs. The $BaTiO_3$ specimen was prepared from a particularly pure grade of titanate made by a chemical precipitation process (Burger and Margolis 1953), the ceramic having been fired at 1500°c. Other ceramic specimens were prepared from commercial grade materials.

---

† A paper by M. McQuarrie and W. R. Buessem (*Bull. Amer. Ceram. Soc.*, 1955, **34**, 402) has just come to the author's notice, in which similar results are described.

For the ageing tests, the specimens were heated to at least 20°c above their Curie point and then plunged into a copper thermostat whose temperature was controlled to within about ± 0·1°c by a mercury contact thermometer. Measurements of capacitance and power factor were taken after five minutes from the time of cooling and at suitable time intervals thereafter.

The capacitance was measured with an accuracy to approximately ± 0·15% at 1 kc/s. The voltage applied was rather less than 0·5 v representing a stress of about 2·5 v cm^{-1}. It was found necessary to apply this voltage continuously to the specimens to obtain consistent readings.

§ 3. Theoretical

3.1. *Permittivity*

The permittivity of ceramic $BaTiO_3$ between room temperature and about 90°c is usually around 1500, while the permittivities in the *a* and *c* directions of the tetragonal single crystal are about 4500 and 300 respectively. For a clamped single crystal these permittivities are expected to be reduced by a factor of up to ten (Devonshire 1951), but no direct experimental data on this are available. In a ceramic the randomly oriented grains and the domains within the grains may be considered as effectively clamping one another, and the average permittivity of randomly oriented free grains (3100) may, therefore, be reduced to as little as 300 when the grains are combined to form a dense ceramic. The observed value of 1500 may, therefore, include a large contribution from the motion of domain walls under the influence of the external applied field. For example, with a field of 1 v cm^{-1} a fraction of only about 10^{-5} of all domai[illegible] need to be reoriented to give a net polarization equivalent to a permittivity of 1500. Kittel (1951) has shown how plausible values of mass, restoring force and total area of domain walls may be combined to account for the observed permittivity.

Consider a diagram such as figure 1, in which the potential energy of a certain area of domain wall is plotted against its position. There is a chance of the wall being caught in the metastable position A when the thermal transformation to the ferroelectric state occurs, and it then requires an activation energy *E* to pass into the stable position B. At A the restoring force for a small applied electric field is relativel[illegible]all, at B it is very much greater. It is therefore suggested that the wall [illegible] contribute to the polarization while at A, but not appreciably at B. The analysis of such a system was carried out by Street and Woolley (1949) in connection with the phenomenon of magnetic viscosity. They

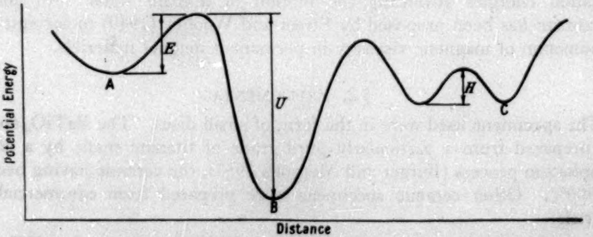


Figure 1. The potential energy of a domain wall as a function of position.

demonstrated that by postulating a very wide distribution of activation energies E a logarithmic increase of magnetic induction with time could be derived. By an analogous argument an equation for the time decrease of permittivity can be arrived at:

$$\epsilon = \epsilon_0 - ST \log t, \qquad \ldots\ldots(1)$$

where ϵ_0 and S are constants, T is the absolute temperature and t the time.

The minimum range over which the activation energies must be spread in order to yield the $\log t$ law over a sufficient time interval can be estimated from the theory. Assuming a natural frequency of oscillation of the domain walls of 10^9 sec^{-1} (Kittel 1951) the range extends from 0·4 to 1·0 ev.

3.2. *Power Factor*

Ceramic $BaTiO_3$ usually has a power factor of 0·01 at frequencies between 50 c/s and 10 Mc/s, and the power factor then rises steeply towards an absorption region in the 1000 Mc/s range (Powles and Jackson 1949). The power factor at the lower frequencies is, undoubtedly, associated with the ferroelectric nature of the ceramic, since it drops to a much lower value above the Curie temperature. It may therefore be associated with 'jumps' of domain walls over low-energy barriers, as shown in figure 1 at C. Fröhlich (1949) has shown that this mechanism leads to a relaxation effect of the Debye type at a frequency given approximately by $v \exp(-H/\boldsymbol{k}T)$. Only a moderate distribution of H values is required to account for the observed spread in relaxation times (0·125 to 0·45 ev if $v = 10^9$ sec^{-1}) and these values of activation energy would suitably complete the distribution indicated above for the ageing of the permittivity. In view of the random character assumed for the potential energy diagram, it is quite possible to have a low barrier on one side and a high potential barrier on the other side of some wells and a domain wall caught in such an arrangement will contribute to the power factor until it passes over the higher barrier to become inactive. The drop in power factor accompanying the ageing of ϵ is, therefore, suitably explained.

§ 4. RESULTS

The ageing of a sample of reasonably pure $BaTiO_3$ is shown in figure 2 for three temperatures. Owing to the phase transitions at about 10°c and −70°c it is not possible to carry out ageing tests at lower temperatures without altering the domain configuration which exists between 30° and 100°c. Hence the prediction of equation (1) that the slope of the (ϵ, $\log t$) curve should vary as the absolute temperature cannot be checked accurately. The values of the slope are 46, 56 (average over the curve) and 70 for the three temperatures 29·5, 65 and 100°c. These increase rather more rapidly than the absolute temperature.

The same sample, after ageing at 29·5°c for three months, was subjected to a d.c. field of 10 kv cm^{-1} for a period of 1 hour. After removing this field the permittivity and power factor were again measured as a function of time, the results being shown in figure 3. A logarithmic law with a low slope holds.

Amongst a large number of different ceramic compositions based on $BaTiO_3$ which were studied, a few gave a more pronounced ageing effect than pure $BaTiO_3$. A sample containing 90% $BaTiO_3$, 10% $SrTiO_3$ and an addition of approximately 0·5% MgF_2 was tested over a period of several years, the plot of permittivity and $\tan\delta$ against time being shown in figure 4. Figure 5 gives plots of $\tan\delta$ against frequency before (curve I) and after (curve II) ageing.

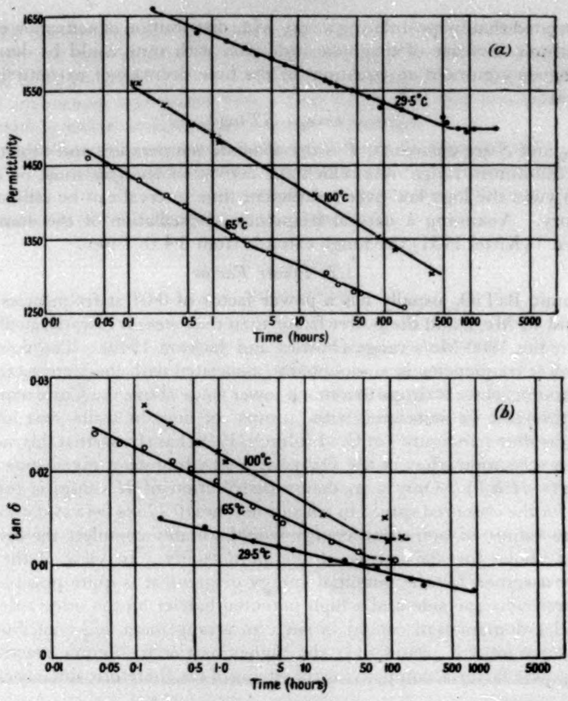


Figure 2. Ageing of the permittivity (*a*) and tan δ (*b*) of the same sample of $BaTiO_3$ at three different temperatures.

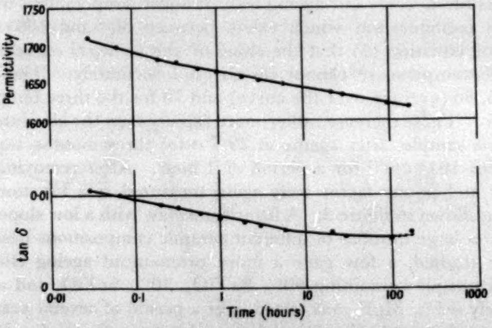


Figure 3. Ageing after the temporary application of a d.c. field of 10 kv cm^{-1} to a well-aged sample of $BaTiO_3$ (temperature 29·5°c),

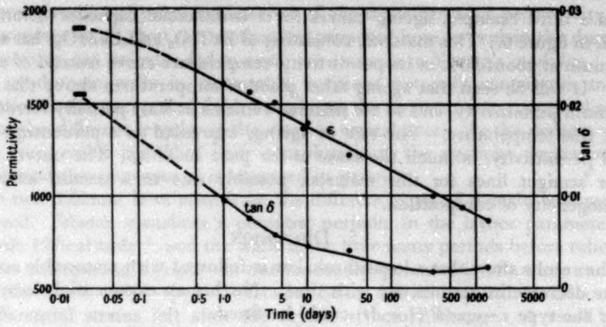


Figure 4. Ageing at room temperature of a specimen containing 90% $BaTiO_3$, 10% $SrTiO_3$ + 0·5% MgF_2.

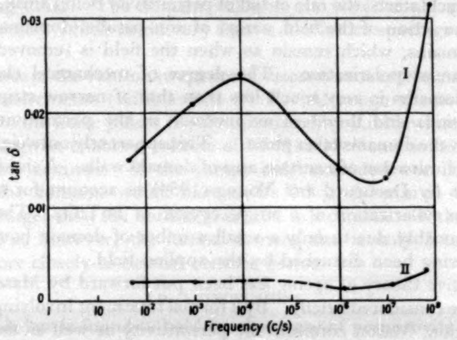


Figure 5. Loss tangent as a function of frequency before and after ageing for a specimen containing 90% $BaTiO_3$, 10% $SrTiO_3$ + 0·5% MgF_2.

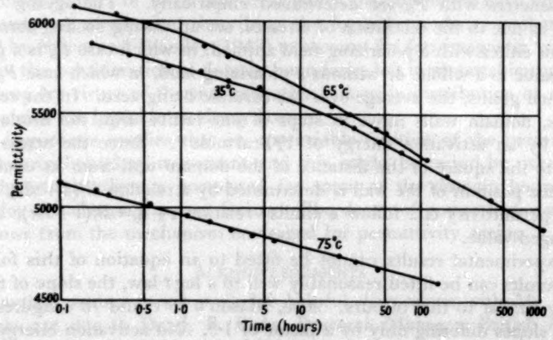


Figure 6. Ageing of a typical capacitor ceramic containing $BaTiO_3$ and $CaZrO_3$.

As a third example, ageing curves for a commercial capacitor ceramic are shown in figure 6. This material, consisting of $BaTiO_3$ and $CaZrO_3$, has a broad maximum at about 50°c in its permittivity–temperature curve instead of a sharp peak. It will be seen that ageing takes place at temperatures above that of the maximum permittivity, and so the material remains at least partially ferroelectric above this temperature. The rate of ageing, expressed as a percentage of the initial permittivity, is much the same as for pure $BaTiO_3$. The curves are no longer straight lines for this material, possibly due to a certain amount of inhomogeneity of composition.

§ 5. DISCUSSION

The results show that a logarithmic law is followed with reasonable accuracy for the decrease in permittivity with time. Neither an exponential decay nor a law of the type $\epsilon - \epsilon_0 = B/(1+At)$ could be fitted.

Referring to figure 3, it will be seen that when $BaTiO_3$ has been polarized by the temporary application of a d.c. field the permittivity is raised slightly and a new ageing cycle starts, the rate of fall of permittivity being small. It is thought that, due to the action of the field, arrays of anti-parallel domains are converted into single domains, which remain so when the field is removed and give the observed remanent polarization. The degree of mechanical clamping of the larger single domains is very much less than that of narrow strips polarized in opposite directions and therefore an increase in the permittivity due to the material within the domains takes place. This apparently outweighs the decrease caused by the elimination of a certain area of domain walls. A similar explanation has been given by Drougard and Young (1954) to account for the increase in permittivity on polarization of a single crystal of $BaTiO_3$. The slow rate of ageing is presumably due to only a small number of domain boundaries of the stable type having been disturbed by the applied field.

A quantitative theory of ageing has been put forward by Mason (1955), and this will now be considered briefly. By a formal treatment involving a generalized enthalpy function, Mason connects the permittivity as well as the piezoelectric and elastic constants with the permanent polarization P_0 of the ceramic, attributing the ageing of all three to a decrease in P_0. The constants connecting these parameters with P_0 are determined empirically. The ageing of P_0 is visualized as due to the relaxation of stresses, set up during cooling through the Curie point either with a polarizing field applied, in which case P_0 is a property of the ceramic as a whole, or without a polarizing field, in which case P_0 applies to individual grains, the average over the ceramic being zero. In the relaxation of stresses, domain walls move in steps of one lattice unit, the motion being governed by an activation energy of 19 kcal mole^{-1}. Since the stress is proportional to the square of the distance of the domain wall from its equilibrium position, the position of the wall is determined by a relation $B/(1+At)$, and the ageing of permittivity etc. follow a similar relation $\epsilon - \epsilon_0 = B/(1+At)$, ϵ_0 being the fully aged value.

Our experimental results cannot be fitted to an equation of this form, but Mason's results can be fitted reasonably well to a $\log t$ law, the slope of the lines being very similar to that of ours. Also, Mason's 25°c and 70°c figures fall on lines with slopes differing only by a factor of 1·3. No activation energy can be derived from the (ϵ, $\log t$) plot.

Mason gives no physical explanation of the dependence of permittivity and elastic constant on P_0, but the mechanism proposed here to explain the decrease in permittivity can be extended to cover the increase in Young's modulus. The motion of domain walls in metastable positions not only contributes to the polarization under an applied field, but also to the strain under an applied stress owing to the different lengths of the *c* and *a* crystal axes. As the domain walls acquire more stable positions, their contribution to the elasticity decreases and Young's modulus increases as is observed in practice.

The two different laws arise from the different types of domain wall motion considered. Mason visualizes a potential periodic in the lattice parameter, of amplitude 19 kcal mole^{-1}, and the walls move over many periods before relieving the stress and coming to rest. In our model the potential is of a more random nature, the more widely-spaced obstacles being due to impurities, dislocations and differential strains left after the crystal structure change. A small scale periodic potential of height about 2 kcal mole^{-1} may well be superimposed on this, the figure of 2 kcal mole^{-1} having been obtained by Merz (1954) from single crystal switching times. The domain walls, in our view, would move in single jumps over distances which could vary from a few lattice spacings to one or two microns.

Some evidence of the volume swept out in discontinuous steps may be obtained from noise measurements (Kibblewhite 1955). Although the average volume corresponding to the noise pulses is only calculated by Kibblewhite for his single crystal, the data given enable a volume of 10^{-10} to $10^{-11}\,\text{cm}^3$ to be estimated for a Ba–Pb titanate ceramic. If the jumps only extended one lattice distance (4×10^{-8} cm), the domains would have to be 0·5 to 0·1 mm in lateral extent, which is much larger than the normal grain size in a ceramic. If the jumps cover a distance of $1\,\mu$, however, the lateral domain size is only 3 to $10\,\mu$, which corresponds more closely to normal ceramic grains.

§ 6. CONCLUSIONS

A logarithmic law is found to hold for the decay of permittivity with time for several types of high permittivity ceramic. A simple exponential decay or a bi-molecular law of the type $B/(1+At)$ does not describe the results over the three or more powers of 10 in time over which measurements have been made.

The simplest mechanism describing the logarithmic decay is that initially proposed by Street and Woolley (1949) to account for ferromagnetic 'viscosity'. It involves the consideration of domain boundaries left in metastable equilibrium which decay by virtue of thermal activation. A wide distribution of activation energies yields the logarithmic decay law.

This hypothesis implies that an appreciable fraction of the permittivity is due to domain boundary movement and this might be tested by other means. A microhysteresis effect is postulated for some of the domain boundaries to account for the observed power factor. The decay of power factor with time then follows from the mechanism envisaged for permittivity ageing.

ACKNOWLEDGMENTS

The author wishes to acknowledge helpful discussions with Mr. R. West and thanks are due to Dr. J. B. Birks, Research Manager, British Dielectric Research Ltd., for permission to publish this work.

References

Burger, F. J. P. J., and Margolis, D. S., 1953, *Brit. Pat.* 691, 257.

Bogoroditskiĭ, N. P., and Verbitskaya, T. N., 1952, *Zh. Tekh. Fiz.*, **22,** 1920.

Devonshire, A. F., 1951, *Phil. Mag.*, **42,** 1065.

Drougard, M. E., and Young, D. R., 1954, *Phys. Rev.*, **94,** 1561.

Fröhlich, H., 1949, *Theory of Dielectrics* (Oxford University Press), p. 78.

Kibblewhite, A. C., 1955, *Proc. Instn Elect. Engrs*, B, **102,** 59.

Kittel, C., 1951, *Phys. Rev.*, **83,** 458.

Marks, B. H., 1948, *Electronics*, **21,** (8), 116.

Mason, W. P., 1955, *J. Acoust. Soc. Amer.*, **27,** 73.

Merz, W. J., 1954, *Phys. Rev.*, **95,** 690.

Novosiltsev, N. V., Khodakov, A. L., and Shulman, M. S., 1952, *Dokl. Akad. Nauk.*, **83,** 829.

Powles, J. G., and Jackson, W., 1949, *Proc. Instn Elect. Engrs*, Pt. III, **96,** 383.

Street, R., and Woolley, J. C., 1949, *Proc. Phys. Soc.* A, **62,** 562.

A Pulse Method for Measurement of Hall Coefficient at Low Temperatures: Some Results on Indium Antimonide

By R. F. BROOM and A. C. ROSE-INNES

Services Electronics Research Laboratory, Baldock, Herts.

MS. received 19*th June* 1956

Abstract. A pulse method has been developed for the measurement of Hall coefficients in high resistivity semiconductors. This technique is specially useful at low temperatures since it greatly reduces heating of the specimens.

Measurements have been made over the range 90°K–1·4°K on several p-type indium antimonide samples of different impurity concentration. From the variation of Hall coefficient with temperature the energy gap between the impurity levels and the valence band has been estimated. The results are consistent with the empirical formula of Pearson and Bardeen (1949) connecting the energy gap with impurity concentration in p-type silicon.

---

§ 1. The Pulse Method

1.1. *Introduction*

We have developed a pulse method for measuring the Hall coefficient of semiconductors. This method is particularly useful at low temperatures in those cases where the resistivity of the specimen becomes very high. In such cases a continuous current of sufficient magnitude to give a reasonable Hall voltage would generate too much heat within the specimen. It is usually necessary to limit the mean power dissipated to, at most, a few milliwatts.

The Hall voltage developed across a specimen is given by $V = RHi/t$ where R is the Hall coefficient of the material, H the magnetic field, i the current through the specimen and t the thickness of the sample in the direction of the field. If P is the power generated in the specimen by the current, $V \propto RH(P\sigma/t)^{1/2}$ where σ is the conductivity of the material. Taking $R = 1/ne$ and $\sigma = ne\mu$, where n, e and μ are the concentration, charge and mobility of the carriers, we obtain

$$V \propto \frac{H}{t^{1/2}}\left(\frac{\mu}{ne}\right)^{1/2} P^{1/2}.$$

When making measurements at low temperatures the power generated must be kept small and, furthermore, conduction will usually be taking place in the impurity band where the mobility is low. Hence we shall only be able to obtain small values of Hall voltage, which it would require very delicate apparatus to measure. With the pulse method described below this difficulty is avoided and, though slightly less accurate than conventional methods, sensitive galvanometers are not required and random errors due to temperature fluctuations during measurement are eliminated. With this technique it is possible to make measurements at liquid helium temperatures on specimens with a Hall mobility as low as 2 cm^2 (volt sec)$^{-1}$.

1.2. *The Pulse System*

A block diagram of the arrangement we have used is shown in figure 1. The pulses of current are derived from a dry battery which is connected to the specimen through a high speed relay driven by a multivibrator. The pulses were of 5 milliseconds duration with a repetition rate of about one per second. This low repetition rate keeps the total heat generated in the specimen small.

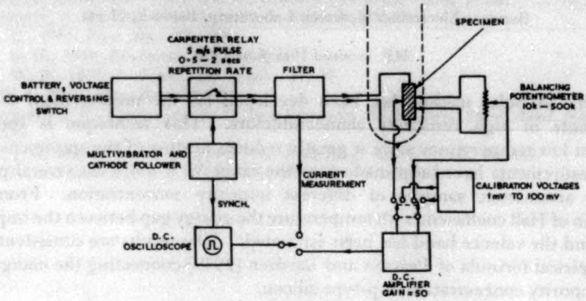

SPECIMEN
CARPENTER RELAY
5 m/s PULSE
0·5 – 2 secs
REPETITION RATE
FILTER
BATTERY, VOLTAGE CONTROL & REVERSING SWITCH
BALANCING POTENTIOMETER
10k – 500k
MULTIVIBRATOR AND CATHODE FOLLOWER
CURRENT MEASUREMENT
CALIBRATION VOLTAGES
1mV TO 100 mV
SYNCH.
D.C. OSCILLOSCOPE
D.C. AMPLIFIER
GAIN = 50

Figure 1. Pulse method for measuring Hall coefficient.

Between the relay and the specimen is a filter, shown in detail in figure 2. This filter serves a dual purpose: firstly, to measure the current and, secondly, to balance the impedance of the system with respect to earth (see below). R_1 is an accurate $1\,\mathrm{k}\Omega$ resistor in series with one lead; the voltage drop across this is measured on a cathode-ray oscilloscope and gives the current flowing through the specimen during the pulse.

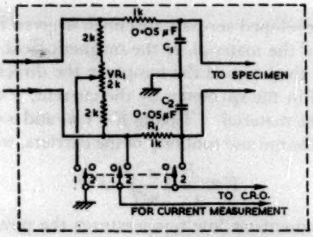

1k
2k
0·05 μF
C_1
VR_1
2k
TO SPECIMEN
C_2
2k
0·05 μF
R_1
1k
1
2
1
2
1
2
TO C.R.O.
FOR CURRENT MEASUREMENT

Figure 2. Details of filter.

The use of a pulse technique raises difficulties not met with in normal d.c. measurements, because the high-frequency components of the pulse can cause coupling between the current and Hall voltage leads. This gives rise to voltage output pulses when no Hall voltage is being generated in the specimen. This coupling is reduced to a minimum when the impedance of the system is balanced with respect to earth. This is done by adjustment of VR_1. The condensers C_1 and C_2, by eliminating the higher frequency components of the pulse, prevent residual unbalance due to the intercapacitance of the leads. The time constant of this RC network is short and the distortion of the pulse it produces is scarcely noticeable on the oscilloscope (see figure 3, Plate).

The contacts to the specimen are arranged as shown in figure 1. Three Hall voltage probes are used: the two on the right-hand side of the specimen are connected to a potentiometer of resistance high compared with the resistance between these probes, allowing one to balance out the ohmic voltage produced because the Hall probes are not exactly opposite each other. The potentiometer and VR_1 (figure 2) are adjusted alternately until zero output voltage is obtained in the absence of a magnetic field. (The two right-hand probes are also used to measure the voltage drop in the material during resistivity experiments.) It is known that the Hall voltage follows the current almost instantaneously (Barlow 1955), and a voltage pulse of the same form as the current pulse is obtained from the voltage probes. This pulse is fed through a d.c. amplifier, whose input is balanced symmetrically with respect to earth, on to the Y plates of the oscilloscope. The use of a d.c. amplifier simplifies calibration (see below). The time base of the oscilloscope is synchronized by a pulse from the multivibrator.

To calibrate the arrangement known d.c. voltages of the same order of magnitude as the Hall voltage can be fed on to the amplifier and the resulting deflection of the oscilloscope trace measured. The usual values of Hall voltage and current through the specimen are taken with the current and magnetic field in either direction. With our d.c. amplifier the minimum Hall voltage measurable is about 250 microvolts with an error of about 10%. In practice we find that we cannot balance the system sufficiently with the filter described above if the resistance of the specimen is more than about 5×10^5 ohms. This gives an upper limit to the resistance of a sample which we can measure.

An important advantage of the pulse method is that measurements are very quickly made. This greatly reduces the tedium of making a large series of measurements. This speed of measurement is also important at low temperatures since the temperature is less likely to vary while the measurement is being carried out.

1.3. *The Cryostat*

The specimen is mounted on a ceramic jig in a copper calorimeter which can be immersed in liquid helium. The jig is supported from a ground cone at the top of the apparatus enabling the specimen to be rotated in the calorimeter. Temperatures below 4·2°K are obtained by boiling liquid helium under reduced pressure in the usual way; temperatures between 4·2°K and 30°K are obtained by means of our desorption system described elsewhere (Rose-Innes and Broom 1956). With our arrangement it is not possible to stabilize temperatures above 30°K and measurements in this region are made as the calorimeter warms up to the temperature of liquid oxygen, but, as this takes about an hour, errors due to temperature drift are quite small.

Below 4·2°K the temperatures are calculated from the vapour pressure of the liquid helium and above 4·2°K a helium gas thermometer, attached to the calorimeter, is used.

The calorimeter is filled to a few millimetres pressure of helium gas to maintain thermal contact with the specimen.

§ 2. Experiments on p-Type InSb

2.1. *Specimens*

Specimens were cut from single crystals grown by the Kyropoulus method from zone refined material. The dimensions were about 1 cm by 2 mm by 1 mm. Current was passed through the specimen by spring contacts pressing against the ends of the specimen which were copper plated. Potential contacts consisted of 0·003 in. platinum wire pressed against the sides of the specimen and welded on by discharging a condenser through the contact.

The samples measured, and the estimated value of the impurity concentration of donors and acceptors, are given in the table. For sample A these concentrations were estimated from the Conwell–Weisskopf formula for impurity scattering using the measured value of the mobility. No information was available on sample B but from our measurements of Hall coefficient and resistivity the impurity concentration must lie somewhere between the limits shown in the table. Samples C and D were so strongly p-type that it was assumed that the concentration of donors N_D was small compared with that of acceptors, so we estimated N_A directly from the Hall coefficient at 90°K. Sample E was polycrystalline, but is included because it is the purest specimen we have obtained so far. The purity was estimated from a knowledge of the segregation coefficients in the zone refining process and the variation of Hall coefficient along the refined ingot.†

p-type InSb Samples Measured

Sample	Form	N_A+N_D (cm^{-3})	N_A (cm^{-3})	H (Oe)
A	Single crystal	$3{\cdot}5\times10^{16}\pm50\%$	$1{\cdot}8\times10^{16}$	750
B	Single crystal	$4–6\times10^{16}$	—	2160
C	Single crystal	7×10^{16}	7×10^{16}	2200
D	Single crystal	8×10^{16}	8×10^{16}	1550
E	Poly-crystal	$4\times10^{15}\pm50\%$	$3{\cdot}6\times10^{15}$	4480 below 7°K 1000 above 7°K

H is the value of magnetic field at which R_H was measured.

2.2. *Hall Coefficient*

Figure 4 shows our results for the specimens listed in the table. It was impossible to carry the results on sample A to below 11°K because the specimen cracked at this temperature. Sample B shows clearly the form associated with impurity band conduction (Hung 1950, Fritzsche and Lark-Horowitz 1955), with approximately the same Hall coefficient at high and low temperature. The more impure samples C and D have the Hall maximum at a higher temperature, and the value of the Hall coefficient is smaller on the low temperature side of the maximum. (This is not very clear on the graph because of the cramped scale at high temperatures but, in fact, there is no change in the Hall coefficient between 90 and 150°K.) We used this high temperature value to estimate the acceptor concentration.

Calculation of the height of the Fermi level (Shockley 1950) shows that these two samples are approaching degeneracy below about 40°K. Sample C was measured with a magnetic field of 2200 oersteds while D was measured at

---

† This method is due to Mr. J. T. Edmond and Dr. F. A. Cunnell of this laboratory. The specimens and the above data were kindly supplied by Dr. F. A. Cunnell.

1550 oersteds and the dependence of R_H upon field may account for the curves crossing over.

The sample E, though purer than the other specimens, was polycrystalline and showed two effects not observed in the other specimens. Firstly, there are two distinct slopes in the Hall coefficient curve above 10°K and, secondly, the

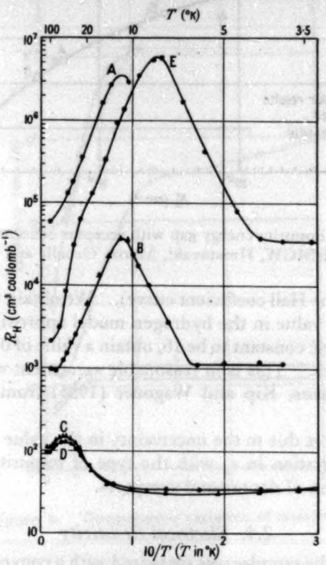


Figure 4. Variation of Hall coefficient with temperature.

variation of Hall coefficient with magnetic field at 90°K was found to be very large, the value of R_H decreasing from 3660 cm³ coulomb⁻¹ at 100 oersteds to 1900 cm³ coulomb⁻¹ at 2000 oersteds. A similar effect is discussed by Willardson, Harman and Beer (1954) for the case of germanium, and they have attributed it to the presence of a small number of 'light' holes (see also Hrostowski, Morin, Geballe and Wheatley 1955).

2.3. *Activation Energy of Acceptors*

We have estimated the activation energy of the acceptors from the slope of the Hall coefficient curve on the high temperature side of the maximum, and in figure 5 this is plotted against the concentration of acceptors N_A. The most reliable values, those from samples A and D, have been used to calculate the constants in the empirical equation used by Pearson and Bardeen (1949) for p-type silicon,

$$\epsilon_A = \epsilon_0 - \alpha N_A{}^{1/3}$$

where ϵ_A is the activation energy at an acceptor concentration N_A, ϵ_0 is the activation energy at infinite dilution and α is a constant. The curve of this equation is drawn in figure 5 and is a reasonable fit with the points from the other three specimens (the activation energy of specimen E was taken from the mean value of the slope of the Hall coefficient curve). We obtain the value $\epsilon_0 = 0{\cdot}010$ ev and, substituting this value in the hydrogen model approximation for impurities and taking the dielectric constant to be 16, obtain a value of 0·19 m_e for the average effective mass for holes. This is in reasonable agreement with the value 0·18 m_e obtained by Dresselhaus, Kip and Wagoner (1955) from cyclotron resonance measurements.

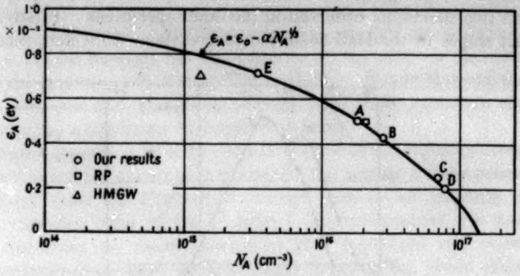


Figure 5. Variation of impurity energy gap with acceptor concentration. RP, Rollin and Petford (1955); HMGW, Hrostowski, Morin, Geballe and Wheatley (1955).

The greatest error is due to the uncertainty in the value of N_A. Other errors could be due to a variation in ϵ_A with the type of impurity present and in the degree of compensation of donors and acceptors.

2.4. *Electrical Resistivity*

The resistivity of the samples was measured with a conventional potentiometer arrangement and the results are shown in figure 6. All specimens exhibit impurity band conduction as shown by the smaller slope of the curves at low temperatures, below the maximum in the Hall curve.

§ 3. CONCLUSIONS

The pulse technique has been found to be a quick and simple method of measuring Hall coefficients in high resistance semiconductors over a wide range of resistivity and temperature. Given uniform specimens with a mobility not less than 2 cm² (volt sec)⁻¹ an error of less than 5% can be obtained.

The activation energy has been calculated from the variation of Hall coefficient with temperature obtained by this method and the results have been found to fit the equation of Pearson and Bardeen connecting activation energy with impurity concentration.

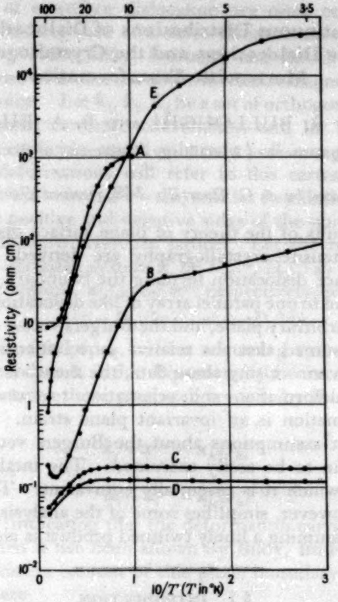


Figure 6. Temperature variation of resistivity.

ACKNOWLEDGMENTS

We wish to express our gratitude to Mr. I. M. Ross, Mr. J. T. Edmond and Dr. F. A. Cunnell for much helpful criticism and advice. Acknowledgment is made to the Admiralty for permission to publish this paper.

REFERENCES

BARLOW, H. E. M., 1955, *Proc. Instn Elect. Engrs*, Pt. B, **102**, 179.
DRESSELHAUS, G., KIP, A. F., and WAGONER, G., 1955, *Phys. Rev.*, **98**, 556.
FRITZSCHE, H., and LARK-HOROWITZ, K., 1955, *Phys. Rev.*, **99**, 400.
HROSTOWSKI, H. J., MORIN, F. J., GEBALLE, T. H., and WHEATLEY, G. H., 1955, *Phys. Rev.*, **100**, 1672.
HUNG, C. S., 1950, *Phys. Rev.*, **79**, 727.
PEARSON, G. L., and BARDEEN, J., 1949, *Phys. Rev.*, **75**, 865.
ROLLIN, B. V., and PETFORD, A. D., 1955, *J. Electronics*, **1**, 171.
ROSE-INNES, A. C., and BROOM, R. F., 1956, *J. Sci. Instrum.*, **33**, 31.
SHOCKLEY, W., 1950, *Electrons and Holes in Semiconductors* (New York : Van Nostrand).
WILLARDSON, R. K., HARMAN, T. C., and BEER, A. C., 1954, *Phys. Rev.*, **96**, 1512.

Continuous Distributions of Dislocations: Surface Dislocations and the Crystallography of Martensitic Transformations

By R. BULLOUGH[†] and B. A. BILBY
Department of Metallurgy, University of Sheffield

Communicated by A. G. Quarrell; *MS. received* 17*th April* 1956

Abstract. Some results of the theory of plane surface dislocations required for the analysis of martensite crystallography are derived. In their subsequent application, the surface dislocation forming the boundary between the phases is assumed to correspond to one parallel array of like dislocations of arbitrary Burgers vector moving in an arbitrary plane, and the Burgers vector and plane are assumed given. It is also assumed that the relative pure lattice strain of the phases is quite general, and given. Using these data, the theory is applied to deduce the habit plane, shape deformation and orientation relationship occurring. The relative shape deformation is an invariant plane strain. The method enables the effect of different assumptions about the Burgers vector, plane of motion, and lattice pure strain to be easily examined. The method is compared with other treatments to which it is essentially equivalent. The use of the surface dislocation theory, however, simplifies some of the analysis. The relation of the method to analyses assuming a finely twinned product is also discussed.

§ 1. Introduction

When an array of dislocation lines in two dimensions is considered as an entity, with its internal structure smoothed out, the idea of a *surface dislocation* arises (Bilby 1955[‡]). An elementary example is the plane array of edge dislocations forming a simple tilt boundary between two identical lattices. The general surface dislocation describes the boundary between two different lattices, and is thus appropriate for the discussion of martensitic transformations, where one phase grows at the expense of the other by the regular motion of an interface between them (Bilby and Christian 1955). Characteristically, in martensitic transformations, the product phase forms as lenticular plates within the parent. The plane of the plates (called the habit plane) bears a definite orientation with respect to each phase, and the material within the plate suffers a shape deformation which is, at least approximately, an invariant plane strain (see for example, the review by Bilby and Christian 1955). In §§ 3, 4 and 5 of this paper the theory of plane surface dislocations is applied to deduce the habit plane, shape deformation and orientation relationship occurring in a given transformation. For this analysis certain results of the theory of surface dislocations are required; these are given in § 2. Finally, in § 6 the method is compared with other treatments of this problem.

[†] Now at Associated Electrical Industries Research Laboratories, Aldermaston, Berks.

[‡] The relevant §§ 4–7 of this paper form a preliminary account of a study carried out in collaboration with Messrs. R. Bullough and E. Smith.

§ 2. PROPERTIES OF SURFACE DISLOCATIONS

The properties of a surface dislocation are most conveniently derived by considering it as a limiting case of a three-dimensional continuous distribution of dislocations (Bilby 1955, Bilby, Bullough and Smith 1955). Its properties are analogous to those of a current sheet in electromagnetic theory, or a vortex sheet in hydrodynamics. Let $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \hat{\mathbf{x}}_3$ be a set of orthogonal unit vectors defining a system of orthogonal cartesian coordinates, and let $\mathbf{x}=x_i\hat{\mathbf{x}}_i$ be any vector. Except where it is explicitly stated otherwise, all components of vectors and representations of deformations will refer to this cartesian system. Consider a plane surface with unit normal $\boldsymbol{\nu}=\nu_i\hat{\mathbf{x}}_i$ and let two lattices with bases $\mathbf{a}_i^{(+)}, \mathbf{a}_i^{(-)}$ be generated on the positive and negative sides of the normal $\boldsymbol{\nu}$ by homogeneous deformation of some third reference lattice. Let the representations of these deformations in the cartesian system $\hat{\mathbf{x}}_i$ be $D_{ij}^{(+)}, D_{ij}^{(-)}$ so that any vector $\mathbf{x}=x_i\hat{\mathbf{x}}_i$ becomes, in the two regions, $\mathbf{x}^{(+)}$ and $\hat{\mathbf{x}}^{(-)}$ respectively, where

$$x_i^{(+)}=D_{ij}^{(+)}x_j, \quad x_i^{(-)}=D_{ij}^{(-)}x_j. \qquad \ldots\ldots(1)$$

Here, as elsewhere in this paper, indices occurring twice in any formula are to be summed over the values 1, 2, 3. Equation (1) means, in fact, that if the bases $\mathbf{a}_i^{(+)}, \mathbf{a}_i^{(-)}$ are chosen to be those produced from $\hat{\mathbf{x}}_i$ by the deformations, then, in the (+) region,

$$\hat{\mathbf{x}}_i \to \mathbf{a}_i^{(+)}=\hat{\mathbf{x}}_k D_{ki}^{(+)} \qquad \ldots\ldots(2)$$

and in the (−) region

$$\hat{\mathbf{x}}_i \to \mathbf{a}_i^{(-)}=\hat{\mathbf{x}}_k D_{ki}^{(-)} \qquad \ldots\ldots(3)$$

the notation $\hat{\mathbf{x}}_i \to \mathbf{a}_i^{(+)}$ indicating that the deformation carries $\hat{\mathbf{x}}_1$ to $\mathbf{a}_1^{(+)}$, $\hat{\mathbf{x}}_2$ to $\mathbf{a}_2^{(+)}$ and $\hat{\mathbf{x}}_3$ to $\mathbf{a}_3^{(+)}$. Then it has been shown by Bilby, Bullough and Smith (Bilby 1955) that the dislocation content of this plane boundary between the lattices is specified by B_{ij}, where

$$B_{ij}=\epsilon_{jkl}\nu_k[E_{il}^{(+)}-E_{il}^{(-)}]. \qquad \ldots\ldots(4)$$

$B_{ij}\hat{\mathbf{x}}_i$ is the resultant Burgers vector of dislocation lines cutting unit length of a line in the boundary perpendicular to the x_j direction.† Here ϵ_{jkl} vanishes unless j, k, l are all different and has the value +1 or −1 according as j, k, l is an even or odd permutation of the numbers 1, 2, 3. The matrices $E_{ij}^{(+)}$ and $E_{ij}^{(-)}$ are the reciprocals of $D_{ij}^{(+)}$ and $D_{ij}^{(-)}$ respectively. In the present work matrices will be represented by capital letters, either in italics with explicit ordered suffixes indicating rows and columns respectively, or in heavy type. The transpose of a matrix will be indicated by a prime, and its reciprocal by the superscript −1. Similarly, vectors will be written as small letters, either in italics with an explicit suffix, or in heavy type as column vectors $\mathbf{n}$ or row vectors $\mathbf{n}'$. Thus the relation between $\mathbf{D}^{(+)}$ and $\mathbf{E}^{(+)}$ may be written $\mathbf{E}^{(+)}\mathbf{D}^{(+)}=\mathbf{D}^{(+)}\mathbf{E}^{(+)}=\mathbf{I}$ where $\mathbf{I}$ is the unit matrix, or $D_{ij}^{(+)}E_{jk}^{(+)}=E_{ij}^{(+)}D_{jk}^{(+)}=\delta_{ik}$ where $\delta_{ik}=1$ if $i=k$ and vanishes otherwise. It is convenient to take the (−) lattice as the reference lattice, and this will be done throughout. Thus $\mathbf{D}^{(-)}=\mathbf{I}$ or $D_{ij}^{(-)}=\delta_{ij}$. If, in fact, the points of the (−) lattice are referred to the cartesian basis $\hat{\mathbf{x}}_i$, the components of (−) lattice vectors, although no longer necessarily rational, are identical with the cartesian coordinates of their end points. The orthonormal basis in the (−) lattice will also be chosen so that it lies along the principal axes of the deformation $D_{ij}^{(+)}$. Thus, in a matrix rotation,

---

† For simplicity of exposition, it is often convenient to talk of the surface dislocation as if it were resolved into discrete sets of dislocation lines. In fact, of course, the analysis deals only with a continuous distribution and its resultant densities for circuits about various vectors drawn in the surface.

$$\mathbf{D}^{(+)} = \mathbf{RP}, \quad \mathbf{E}^{(+)} = \mathbf{P}^{-1}\mathbf{R}^{-1} \qquad \ldots\ldots(5)$$

where

$$\mathbf{P} = \begin{bmatrix} e_1 & 0 & 0 \\ 0 & e_2 & 0 \\ 0 & 0 & e_3 \end{bmatrix} \qquad \ldots(6)$$

is a diagonal matrix, and $\mathbf{R}$ is an orthogonal matrix representing a general rigid body rotation. The reciprocal $\mathbf{R}^{-1}$ of $\mathbf{R}$ is equal to its transpose $\mathbf{R}'$.

The relation (4) may be expressed in a slightly different form. Let $\mathbf{p}$ be any unit vector in the boundary plane. Then a unit vector $\mathbf{w}$ in the boundary and normal to $\mathbf{p}$ has components $w_i = \epsilon_{ikl}\nu_k p_l$ and the resultant Burgers vector of dislocation lines cut by $\mathbf{p}$ is $B_{ij}w_j\hat{\mathbf{x}}_i$. Then

$$\begin{aligned} b_i &= B_{ij}w_j \\ &= \epsilon_{jkl}\nu_k[E_{il}^{(+)} - E_{il}^{(-)}]\epsilon_{jmn}\nu_m p_n \\ &= (E_{il}^{(+)} - \delta_{il})p_l \end{aligned} \qquad \ldots\ldots(7)$$

since $\epsilon_{jkl}\epsilon_{jmn} = (\delta_{km}\delta_{ln} - \delta_{kn}\delta_{lm})$ and $\nu_n p_n = 0$.

As the martensitic transformation occurs the product grows at the expence of the parent by the sideways motion of the surface dislocation forming the boundary between them. The surface dislocation moves into the parent leaving the product in its wake. Thus the surface dislocation must be one which can readily move without changing its form. The mobility of dislocation boundaries has been discussed by Read (1954) and Smith (private communication); a full treatment is complicated and will not be given here. It is clear, however, that if a surface dislocation corresponding to one parallel array of like dislocations moves so that the individual motion of each dislocation is glissile, it will be readily mobile. Let the dislocations lie parallel to the unit vector $n_i\hat{\mathbf{x}}_i$ in the reference lattice and have Burgers vectors parallel to the unit vector $l_i\hat{\mathbf{x}}_i$ of this lattice; then they move in the plane of unit normal $m_i\hat{\mathbf{x}}_i$ in this lattice so that

$$l_i m_i = n_i m_i = 0. \qquad \ldots\ldots(8)$$

Such a surface dislocation is called a *simple glissile surface* dislocation, and it will be assumed here that the boundary between the two transforming phases is of this type. Should the theory of the crystallography of martensitic transformations thus developed prove inadequate, more complicated assumptions may be required.

Consider now the motion of such a simple glissile surface dislocation and, for simplicity, let it move *into* the reference lattice (the (−) lattice) in which the quantities l_i, m_i and n_i are defined. In discussing a martensitic transformation in this way, therefore, the parent phase must be taken as the reference lattice, since it is this phase which is steadily consumed. For such a surface dislocation, if different unit vectors $\mathbf{p}^{(k)}$ are chosen in the boundary the corresponding Burgers vectors $\mathbf{b}^{(k)}$ given by (7) may be written $t^{(k)}\mathbf{l}$ and (7) becomes

$$t^{(k)}\mathbf{l} = (\mathbf{P}^{-1}\mathbf{R}^{-1} - \mathbf{I})\mathbf{p}^{(k)}. \qquad \ldots\ldots(9)$$

Clearly $t^{(k)} = 0$ when $\mathbf{p}^{(k)}$ is parallel to $\mathbf{n}$.

In the region swept out by the surface dislocation, the material undergoes a change of shape due to the slip occurring, and at the same time undergoes a further shape change because, in this region, the $(-)$ lattice deforms into the $(+)$ lattice. Thus, as has been emphasized (Bilby 1955), not only is there a discontinuity of *lattice* strain across a surface disiocation, but also a discontinuity of *shape* strain (Bilby and Christian 1955). The shape strain due to slip is readily calculated. For consider any vector $\mathbf{u}$ scribed in the parent phase. Write $\mathbf{u}$ in the form $a\mathbf{l}+b\mathbf{p}^{(k)}$, where $\mathbf{p}^{(k)}$ is a unit vector such that $\boldsymbol{\nu}'\mathbf{p}^{(k)}=0$. Then $a=(\boldsymbol{\nu}'\mathbf{u})/(\boldsymbol{\nu}'\mathbf{l})$. As the surface dislocation passes, $a\mathbf{l}$ is unchanged, but the ends of $b\mathbf{p}^{(k)}$ suffer a relative displacement equal to the resultant Burgers vector of the dislocation distribution cut by $b\mathbf{p}^{(k)}$. Thus, due to slip, the vector $b\mathbf{p}^{(k)}$ becomes $b(\mathbf{p}^{(k)}+t^{(k)}\mathbf{l})$ and $\mathbf{u}$ becomes $\mathbf{v}$ where $\mathbf{v}=\mathbf{u}+bt^{(k)}\mathbf{l}$. If $\mathbf{m}'\mathbf{u}=0$, $\mathbf{p}^{(k)}=\mathbf{n}$, $t^{(k)}=0$ and $\mathbf{v}=\mathbf{u}$. If $\mathbf{u}$ increases by a factor c, so also does the increment $\mathbf{v}-\mathbf{u}$. Thus this increment is also proportional to $\mathbf{m}'\mathbf{u}$ and $\mathbf{v}=\mathbf{S}\mathbf{u}$ where

$$\mathbf{S}-\mathbf{I}=g\begin{bmatrix} m_1l_1, & m_2l_1, & m_3l_1 \\ m_1l_2, & m_2l_2, & m_3l_2 \\ m_1l_3, & m_2l_3, & m_3l_3 \end{bmatrix} \qquad \ldots\ldots(10)$$

and

$$t^{(k)}=g\mathbf{m}'\mathbf{p}^{(k)}. \qquad \ldots\ldots(11)$$

The slip shape strain is thus a simple shear of amount g on the plane $\mathbf{m}$ in the $\mathbf{l}$ direction. As the surface dislocation passes, however, all vectors originally scribed in the reference lattice also suffer the lattice deformation $\mathbf{RP}$. Thus finally the scribed vector $\mathbf{u}$ becomes $\mathbf{F}\mathbf{u}$ where

$$\mathbf{F}=\mathbf{RPS}. \qquad \ldots\ldots(12)$$

$\mathbf{F}$ specifies the total shape deformation associated with the motion of the surface dislocation.

Now let $\mathbf{u}$ be any unit vector $\mathbf{p}^{(k)}$ lying in a plane parallel to that of the surface dislocation ($a=0$, $b=1$). Then, due to the slip alone, it becomes $\mathbf{S}\mathbf{p}^{(k)}$ where

$$\begin{aligned}\mathbf{S}\mathbf{p}^{(k)}&=\mathbf{p}^{(k)}+t^{(k)}\mathbf{l}\\ &=\mathbf{P}^{-1}\mathbf{R}^{-1}\mathbf{p}^{(k)}\end{aligned} \qquad \ldots\ldots(13)$$

by equation (9). Hence

$$\mathbf{RPS}\mathbf{p}^{(k)}=\mathbf{p}^{(k)} \qquad \ldots\ldots(14)$$

and so the shape deformation (12) applied to any vector lying in the plane whose normal is $\boldsymbol{\nu}$ leaves this vector invariant. Thus $\mathbf{F}$ is an invariant plane strain with the plane whose normal is $\boldsymbol{\nu}$ as its invariant plane. This result may be stated in a slightly different way. If two lattices deform into each other by the (shape and lattice) deformation $\mathbf{RP}$, and are separated by a simple glissile surface dislocation, then (9) gives the distribution of dislocations that must be inserted between them in order that they may meet on a plane of contact without far-reaching stress. In particular, if the relative (lattice) deformation $\mathbf{D}$ of the two lattices is an invariant plane strain on the plane whose normal is $\boldsymbol{\nu}$, then $B_{ij}=0$, (Bilby and Smith, unpublished). For then $E_{ij}^{(+)}=\delta_{ij}+v_i\nu_j$ where v_i is the displacement of all points at unit distance from the invariant plane. Therefore

$$\begin{aligned}B_{ij}&=\epsilon_{jkl}\nu_k(\delta_{il}+v_i\nu_l-\delta_{il})\\ &=0.\end{aligned} \qquad \ldots\ldots(15)$$

Thus, no distribution of dislocations is required, because there is exact matching of the two lattices on this plane.

§ 3. THE HABIT PLANE OF THE TRANSFORMATION

It is first necessary to determine the vector $\mathbf{n}$, lying parallel to the dislocation lines. From (9), with $\mathbf{p}^{(k)}=\mathbf{n}$, $t^{(k)}=0$,

$$\mathbf{R}'\mathbf{n}=\mathbf{P}\mathbf{n}, \qquad \mathbf{n}'\mathbf{R}=\mathbf{n}'\mathbf{P}' \qquad \ldots\ldots(16)$$

and hence, since $\mathbf{P}$ is diagonal,

$$\mathbf{n}'(\mathbf{I}-\mathbf{P}^2)\mathbf{n}=0 \qquad \ldots\ldots(17)$$

or

$$n_1^2(1-e_1^2)+n_2^2(1-e_2^2)+n_3^2(1-e_3^2)=0. \qquad \ldots\ldots(18)$$

Since $\mathbf{n}$ lies in $\mathbf{m}$,

$$n_1m_1+n_2m_2+n_3m_3=0 \qquad \ldots\ldots(19)$$

and, since $|\mathbf{n}|=1$,

$$n_1^2+n_2^2+n_3^2=1. \qquad \ldots\ldots(20)$$

Provided that $\mathbf{P}$ and $\mathbf{m}$ satisfy certain conditions (18), (19) and (20) determine four real vectors $\pm\mathbf{n}^{(1)}$ and $\pm\mathbf{n}^{(2)}$. The first condition is that the quantities $1-e_1^2$, $1-e_2^2$, $1-e_3^2$ should not all have the same sign. This means that $\mathbf{P}$ must leave some cone of lines unchanged in length, and so must form an invariant line strain if combined with a suitable rotation. It has already been noted that, with the intuitive choices for $\mathbf{P}$, all known martensitic transformations apparently conform to this requirement (Bilby and Christian 1955), and it is possible that the condition may be one restricting the choice of $\mathbf{P}$. The second condition is that

$$m_1^2(1-e_2^2)(1-e_3^2)+m_2^2(1-e_3^2)(1-e_1^2)+m_3^2(1-e_1^2)(1-e_2^2)\leqslant 0. \qquad \ldots\ldots(21)$$

This is the condition that the plane (19) should intersect the cone (18) in two real lines (distinct or coincident). Given $\mathbf{P}$, it restricts the choice of the slip plane $\mathbf{m}$. $\boldsymbol{\nu}$ may now be determined as follows.† One at least of ν_1, ν_2, ν_3 is not zero. Assume that $\nu_3\neq 0$ and write $q_i=\nu_i/\nu_3$, $i=1, 2, 3$, so that $\boldsymbol{\nu}=\nu_3(q_1, q_2, 1)$. If, in fact $\nu_3=0$, it is necessary to define alternative q_i by dividing by a non-zero ν_i. The resulting modification in the analysis is obtained by a suitable permutation of the components of the vectors $\mathbf{p}^{(1)}$, $\mathbf{p}^{(2)}$ chosen below.

Multiplying the equation (14) by its transpose gives

$$\mathbf{p}^{(k)\prime}\mathbf{S}'\mathbf{P}^2\mathbf{S}\mathbf{p}^{(k)}=\mathbf{p}^{(k)\prime}\mathbf{p}^{(k)} \qquad \ldots\ldots(22)$$

where, explicitly, $\mathbf{S}'\mathbf{P}^2\mathbf{S}=\mathbf{A}$ and,

$$A_{ij}=\sum_{k=1}^{3}\{e_k^2(\delta_{ik}+m_il_kg)(\delta_{kj}+m_jl_kg)\}. \qquad \ldots\ldots(23)$$

In (21), choose $\mathbf{p}^{(k)}=\mathbf{p}^{(1)}=\nu_3[0,\ 1,\ -q_2]$. Then

$$h^2P+2h(l_3e_3^2q_2-l_2e_2^2)+(e_3^2-1)q_2^2+e_2^2-1=0 \qquad \ldots\ldots(24)$$

where

$$P=\sum_{j=1}^{3}l_j^2e_j^2 \qquad \ldots\ldots(25)$$

and

$$h=g(m_3q_2-m_2). \qquad \ldots\ldots(26)$$

Multiply (9) on the left by $\mathbf{n}'\mathbf{P}^2$, then

$$t^{(k)}\mathbf{n}'\mathbf{P}^2\mathbf{l}=\mathbf{n}'(\mathbf{P}\mathbf{R}'-\mathbf{P}^2)\mathbf{p}^{(k)}.$$

But, from (16)

$$\mathbf{n}'\mathbf{P}\mathbf{R}'=\mathbf{n}'\mathbf{R}\mathbf{R}'=\mathbf{n}'$$

---

†See note added in proof at end of paper.

thus

$$t^{(k)}\mathbf{n}'\mathbf{P}^2\mathbf{l}=\mathbf{n}'(\mathbf{I}-\mathbf{P}^2)\mathbf{p}^{(k)}. \qquad \ldots\ldots(27)$$

Now take $\mathbf{p}^{(k)}=\mathbf{p}^{(2)}=\nu_3[q_2, -q_1, 0]$; then by (11), $t^{(2)}=g(\nu_2m_1-\nu_1m_2)$. Also, since $\mathbf{n}'\boldsymbol{\nu}=0$, q_1 may be eliminated by using the relation

$$n_1q_1+n_2q_2+n_3=0. \qquad \ldots\ldots(28)$$

Substitution of $\mathbf{p}^{(k)}=\mathbf{p}^{(2)}$ in (27) then gives

$$hQ=q_2n_3(1-e_3{}^2)-n_2(1-e_2{}^2) \qquad \ldots\ldots(29)$$

where

$$Q=\sum_{k=1}^{3} l_kn_ke_k{}^2. \qquad \ldots\ldots(30)$$

Elimination of h between (29) and (24) leads to a quadratic equation for q_2:

$$L_3q_2{}^2+Mq_2+L_2=0 \qquad \ldots\ldots(31)$$

where

$$L_k=N_k\{N_kP+2T_kQ-(Q^2/n_k)\} \qquad \ldots\ldots(32)$$

$$-\tfrac{1}{2}M=N_2N_3P+(N_3T_2+N_2T_3)Q \qquad \ldots\ldots(33)$$

and

$$\left.\begin{aligned} N_k&=n_k(1-e_k{}^2) \qquad k=2,3\\ T_k&=l_ke_k{}^2 \qquad k=2,3\end{aligned}\right\}\ldots\ldots(34)$$

where the summation convention is *not* used in (32) and (34). The habit plane normal is determined by (28) and (31). If $-\mathbf{n}$ is substituted for $\mathbf{n}$ in (30) and (32)–(34), the constants L_k and M are unchanged. Hence, of the four real vectors $\pm\mathbf{n}^{(1)}$, $\pm\mathbf{n}^{(2)}$ determined by (18)–(20) only two, say $\mathbf{n}^{(1)}$, $\mathbf{n}^{(2)}$ lead to distinct values for $\boldsymbol{\nu}$. Each of these, however, leads to two values for $\boldsymbol{\nu}$ because either root of (31) may be taken for q_2. Thus, in general, if $\mathbf{P}$, $\mathbf{l}$ and $\mathbf{m}$ are given, there are four distinct values for $\boldsymbol{\nu}$, the habit plane normal (the possible choices for ν_3 merely reverse the sense of these four vectors). Of course, for special $\mathbf{P}$, $\mathbf{l}$ and $\mathbf{m}$, there may be degeneracy or crystallographic equivalence of some of these solutions.

Similarly, if $\mathbf{l}$ changes sign, L_k and M are unaffected, so that the $\boldsymbol{\nu}$ values are also unchanged. The slip shear, however, changes sign, as is evident from the expression for g, which follows at once from (26) and (29):

$$g=\frac{q_2n_3(1-e_3{}^2)-n_2(1-e_2{}^2)}{Q(m_3q_2-m_2)}. \qquad \ldots\ldots(35)$$

If $\mathbf{m}$ changes sign, $\mathbf{n}$ and $\boldsymbol{\nu}$ are unaffected, but g changes sign. Of course, if $\mathbf{l}$ and $\mathbf{m}$ both change sign, g does not. The sign of g is unaffected if $\mathbf{n}$ changes sign, so that, given $\mathbf{l}$ and $\mathbf{m}$, there is one g for each habit plane $\boldsymbol{\nu}$.

§ 4. THE TOTAL SHAPE DEFORMATION

As has been shown, the total shape deformation $\mathbf{F}$, given by (13), is an invariant plane strain. Since $\mathbf{S}$ is given by (10) and (35), only $\mathbf{R}$ is unknown. It may be found as follows.† For any unit vector scribed in the habit plane, (14) gives

$$\mathbf{PSp}^{(k)}=\mathbf{R}'\mathbf{p}^{(k)}. \qquad \ldots\ldots(36)$$

That is, the application of $\mathbf{PS}$ merely causes the vector to undergo a rotation exactly opposite to that produced by applying $\mathbf{R}$ to it. Now it is easily shown that if $\mathbf{R}$ is a right-hand rotation through an angle θ about a unit vector $\boldsymbol{\alpha}$, and $\mathbf{r}^{(1)}$, $\mathbf{r}^{(2)}$, are any two vectors,

---

† A similar method has also been used in a recent paper (Lieberman, Wechsler and Read 1955).

$$[\tan(\theta/2)]\boldsymbol{\alpha}=\frac{(\mathbf{r}^{(2)}-\mathbf{R}'\mathbf{r}^{(2)})\times(\mathbf{r}^{(1)}-\mathbf{R}'\mathbf{r}^{(1)})}{(\mathbf{r}^{(2)}-\mathbf{R}'\mathbf{r}^{(2)}).(\mathbf{r}^{(1)}+\mathbf{R}'\mathbf{r}^{(1)})}. \qquad \ldots\ldots(37)$$

Hence θ and $\boldsymbol{\alpha}$ may be found by choosing any two vectors in the habit plane and using (36). Convenient choices are $\mathbf{r}^{(1)}=\mathbf{n}$, for which, by (16), $\mathbf{R}'\mathbf{r}^{(1)}=\mathbf{P}\mathbf{n}$; and $\mathbf{r}^{(2)}=\boldsymbol{\nu}\times\mathbf{n}$ for which $\mathbf{S}\mathbf{r}^{(2)}=\mathbf{r}^{(2)}+g(\mathbf{m}'\mathbf{r}^{(2)})\mathbf{l}$, by (13) and (11). In terms of $\boldsymbol{\alpha}$ and θ the matrix elements of $\mathbf{R}$ are

$$R_{ij}=\delta_{ij}\cos\theta+\alpha_i\alpha_j(1-\cos\theta)-\epsilon_{ijk}\alpha_k\sin\theta. \qquad \ldots\ldots(38)$$

When $\mathbf{F}$ is known quantities associated with the shape deformation and which can be determined experimentally are readily calculated. A convenient measure of the magnitude of $\mathbf{F}$ is the angle γ between $\boldsymbol{\nu}$ and $\mathbf{F}\boldsymbol{\nu}$. This is given by

$$\cos\gamma=(\boldsymbol{\nu}'\mathbf{F}\boldsymbol{\nu})/|\mathbf{F}\boldsymbol{\nu}|. \qquad \ldots\ldots(39)$$

The matrix $\mathbf{R}$ determined by (37) is unique (if $0\leqslant\theta\leqslant2\pi$), and does not depend on the sign of $\boldsymbol{\nu}$ or $\mathbf{n}$. Thus, since for given $\boldsymbol{\nu}$ there is a unique g, there is a unique $\mathbf{S}$ by (10). Thus, given $\mathbf{P}$, $\mathbf{l}$ and $\mathbf{m}$, for each habit plane $\boldsymbol{\nu}$ there is one total shape deformation $\mathbf{F}$.

§ 5. THE ORIENTATION RELATIONSHIP

In this section a convenient notation for describing changes of basis in general lattices is used. All quantities associated with the same basis are denoted by the same letter. The basis $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ of the direct lattice is written as a column vector and enclosed in square brackets $[\mathbf{a}]$; that of the reciprocal lattice, $\mathbf{a}^1$, $\mathbf{a}^2$, $\mathbf{a}^3$, as a row vector in round brackets $(\mathbf{a})$. Components in the direct lattice are written as row vectors $\mathbf{a}'$, and of the reciprocal lattice as column vectors $\mathbf{a}$. Matrices connecting two bases $\mathbf{a}$, $\mathbf{b}$ are written $(\mathbf{A}|\mathbf{B})$, with the reciprocal $(\mathbf{B}|\mathbf{A})$. The standard transformation formulae then follow merely by keeping the same letters adjacent. Thus if $[\mathbf{b}]=(\mathbf{B}|\mathbf{A})[\mathbf{a}]$ then $(\mathbf{b})=(\mathbf{a})(\mathbf{A}|\mathbf{B})$, $\mathbf{b}'=\mathbf{a}'(\mathbf{A}|\mathbf{B})$ and $\mathbf{b}=(\mathbf{B}|\mathbf{A})\mathbf{a}$.

The reference bases used in this analysis are the orthogonal unit set $\hat{\mathbf{x}}_i$ chosen in the parent phase and the set $\mathbf{a}_i^{(+)}$ in the product generated from $\hat{\mathbf{x}}_i$ by the lattice deformation $D_{ij}^{(+)}$ of equation (2). In the notation of this section, and dropping the + superscript,

$$[\hat{\mathbf{x}}]\rightarrow[\mathbf{a}]=(\mathbf{A}|\mathbf{X})[\hat{\mathbf{x}}]. \qquad \ldots\ldots(40)$$

The rows of $(\mathbf{A}|\mathbf{X})$ are the components of the basis vectors $[\mathbf{a}]$ of the product in their final orientation with respect to the basis $[\hat{\mathbf{x}}]$ in the parent. Any vector of the direct lattice $\mathbf{a}'[\mathbf{a}]$ has components in the $[\hat{\mathbf{x}}]$ system given by

$$\mathbf{x}'=\mathbf{a}'(\mathbf{A}|\mathbf{X}) \qquad \ldots\ldots(41)$$

where

$$(A|X)_{ik}=D_{ki} \quad\text{or}\quad (\mathbf{A}|\mathbf{X})=\mathbf{D}'=\mathbf{P}\mathbf{R}'. \qquad \ldots\ldots(42)$$

Also, $\mathbf{x}'[\hat{\mathbf{x}}]$ has components $\mathbf{a}'$ in the $[\mathbf{a}]$ system where

$$\mathbf{a}'=\mathbf{x}'(\mathbf{X}|\mathbf{A}). \qquad \ldots\ldots(43)$$

(41) and (43) specify the orientation relationship for directions; thus the angle

between $\mathbf{x}'[\hat{\mathbf{x}}]$ and $\mathbf{a}'[\mathbf{a}]$ is that between $\mathbf{x}'[\hat{\mathbf{x}}]$ and $\mathbf{a}'(\mathbf{A}|\mathbf{X})[\hat{\mathbf{x}}]$. Because $[\hat{\mathbf{x}}]$ is an orthogonal unit basis, this angle is

$$\cos^{-1}\{(\delta_{ij}x^i a^k(A|X)_k{}^j)/(\delta_{ij}x^i x^j)(\delta_{ij}a^k(A|X)_k{}^i a^m(A|X)_m{}^j)\}.$$

Here explicit indices have been introduced, using superscripts for columns and subscripts for rows, and δ_{ij} is the metric of $[\hat{\mathbf{x}}]$. For planes, the relations analogous to (41) and (43) are

$$\mathbf{x}=(\mathbf{X}|\mathbf{A})\mathbf{a} \qquad \ldots\ldots(44) \qquad\qquad \mathbf{a}=(\mathbf{A}|\mathbf{X})\mathbf{x}. \qquad \ldots\ldots(45)$$

These give the components $\mathbf{x}$ in the reciprocal lattice $(\hat{\mathbf{x}})$ of the parent, of the vector $(\mathbf{a})\mathbf{a}$ of the reciprocal lattice of the product, and vice versa. Usually, the bases $[\hat{\mathbf{x}}]$ and $[\mathbf{a}]$ will not be the conventional ones for the parent and product, since $[\hat{\mathbf{x}}]$ is chosen so that $\mathbf{P}$ is diagonal and $[\mathbf{a}]$ is that generated from $[\hat{\mathbf{x}}]$ by $\mathbf{PR}'$. (The basis $[\mathbf{a}^*]$ generated from $[\hat{\mathbf{x}}]$ by $\mathbf{P}$ alone is that usually derived intuitively, and relations similar to (42)–(45) but with $(\mathbf{A}|\mathbf{X})=\mathbf{P}$ then may be used to specify the *correspondence* of vectors and planes). Suppose that the conventional reference bases for parent and product are $[\mathbf{c}]$ and $[\mathbf{d}]$ respectively, and that

$$[\mathbf{c}]=(\mathbf{C}|\mathbf{X})[\hat{\mathbf{x}}] \qquad \ldots\ldots(46) \qquad\qquad [\mathbf{d}]=(\mathbf{D}|\mathbf{A})[\mathbf{a}] \qquad \ldots\ldots(47)$$

then the relations analogous to (42)–(45) specifying the orientation relationship between vectors and planes indexed with respect to the bases $[\mathbf{c}]$ and $[\mathbf{d}]$ are as follows. The product direct lattice vector $\mathbf{d}'[\mathbf{d}]$ has components $\mathbf{c}'$ with respect to the parent basis $[\mathbf{c}]$ where

$$\mathbf{c}'=\mathbf{d}'(\mathbf{D}|\mathbf{A})(\mathbf{A}|\mathbf{X})(\mathbf{X}|\mathbf{C}) \qquad \ldots\ldots(48)$$

and similarly,

$$\mathbf{d}'=\mathbf{c}'(\mathbf{C}|\mathbf{X})(\mathbf{X}|\mathbf{A})(\mathbf{A}|\mathbf{D}) \qquad \ldots\ldots(49)$$

$$\mathbf{c}=(\mathbf{C}|\mathbf{X})(\mathbf{X}|\mathbf{A})(\mathbf{A}|\mathbf{D})\mathbf{d} \qquad \ldots\ldots(50)$$

$$\mathbf{d}=(\mathbf{D}|\mathbf{A})(\mathbf{A}|\mathbf{X})(\mathbf{X}|\mathbf{C})\mathbf{c}. \qquad \ldots\ldots(51)$$

In calculating the angles between vectors and planes in the $[\mathbf{c}]$ and $[\mathbf{d}]$ representations, the appropriate metrics c_{ij}, c^{ij}, d_{ij}, d^{ij} must, of course, be used.

§ 6. COMPARISON WITH OTHER ANALYSES†

Recent analyses of the crystallography of martensitic transformations have been reviewed by Bilby and Christian (1955). The use of the theory of surface dislocations focuses attention on the structure of the interface. The present treatment is thus most closely related to the matching method (Frank 1953, Bilby and Frank, unpublished‡). It is also essentially equivalent to treatments which resolve the deformation connecting the initial and final states into (hypothetical) sequences (Bowles and Mackenzie 1954 a, b and Mackenzie and Bowles 1954, to be referred to as BM, and Wechsler, Lieberman and Read 1954 and Lieberman, Wechsler and Read 1955, to be referred to as WLR). In the present analysis there is no shape deformation of the interface. This is also the assumption of the WLR theory. On the other hand, in the BM theory a uniform shape dilatation of the interface is allowed, and provides an extra degree of freedom giving the theory an additional flexibility. Similar flexibility could be introduced into the surface dislocation treatment by relaxing the requirement that the interface be of the simple glissile type. The BM hypothesis is to restrict the vectors **l** and **m** (to correspond to a twinning direction and a twinning plane of the product) and to choose the dilatation to give the best agreement between the observed and calculated habit planes. The present treatment affords a convenient method of varying **l** and **m** while retaining the interface with no shape deformation (as also does the prism matching method). This procedure is facilitated by the simple parametric solution obtained for the habit plane in the general case. This simplicity is partly due to the fact that although all the deformations are occurring in the interface as it moves, the surface dislocation method here adopted, where the parent is the reference lattice, leads naturally to these deformations being taken formally in a certain order. This order is different from that used in the other theories, and considerably simplifies the analysis. Of course, by changing the reference lattice, the formal order can be altered.

---

† Comparison of recent papers by J. W. Christian (*J. Inst. Met.*, 1956, **84**, 386) which have appeared since the present work was submitted will be given elsewhere.

‡ Reported in part at the Symposium on Martensite, Congress of the International Union of Crystallography, Paris 1954.

At the present time no satisfactory explanation is afforded by any of the theories of why the habit plane sometimes varies with small composition changes of the phases (for example in the iron–carbon alloys) or with the mode of formation (by cold-work, or by cooling, for example). A number of criteria suggest themselves as a possible basis for this selection, and can be quickly examined with the present solution. Thus it is to be expected, as emphasized in the prism matching method (see Bilby and Christian 1955), that both **l**, and the vector **Pl** to which it corresponds in the product, will be vectors associated with the modes of slip or twinning in both phases, although not perhaps the normal ones. Similar remarks apply to the plane **m**. Again the magnitude of g (equation (35)) measures the density of dislocations in the interface, and so transformation habits with small g may be favoured. Similarly, habits with small total shape deformations, for example as specified by the γ of (39) may be preferred. Another possibility, of course, is that the initial stages of the transformation may determine the habit; to examine this, however, a more detailed model of the nucleation stage than is yet available is required.

Finally, there is evidence that in some transformations the adjustment of shape deformation is achieved by alternating twinning on a fine scale in the product. This situation may be brought readily within the scope of the present treatment. It is only necessary to determine, in any given case, the equivalent slip direction **l**. Let the material on the (+) side of the surface dislocation consist of coherent alternating regions H and K, which are generated from that on the (−) side by the shape deformations $\mathbf{D}_{\mathrm{H}}$, $\mathbf{D}_{\mathrm{K}}$ respectively. Thus, in the two regions a vector $\mathbf{x} = x_i\hat{\mathbf{x}}_i$ of the cartesian system $\hat{\mathbf{x}}_i$ becomes $\mathbf{x}_{\mathrm{H}}$, $\mathbf{x}_{\mathrm{K}}$:

$$\mathbf{x} \to \mathbf{x}_{\mathrm{H}} = \mathbf{D}_{\mathrm{H}}\mathbf{x} \qquad \text{......(52)}$$

$$\mathbf{x} \to \mathbf{x}_{\mathrm{K}} = \mathbf{D}_{\mathrm{K}}\mathbf{x} \qquad \text{......(53)}$$

and the relative deformation expressed in the $\hat{\mathbf{x}}_i$ system is

$$\mathbf{x}_{\mathrm{H}} \to \mathbf{x}_{\mathrm{K}} = \mathbf{D}_{\mathrm{K}}\mathbf{D}_{\mathrm{H}}^{-1}\mathbf{x}_{\mathrm{H}}. \qquad \text{......(54)}$$

The deformations $\mathbf{D}_{\mathrm{H}}$ and $\mathbf{D}_{\mathrm{K}}$ may be resolved as follows:

$$\mathbf{D}_{\mathrm{H}} = \mathbf{R}_{\mathrm{H}}\mathbf{Q}_{\mathrm{H}}, \qquad \mathbf{D}_{\mathrm{K}} = \mathbf{R}_{\mathrm{K}}\mathbf{Q}_{\mathrm{K}} \qquad \text{......(55)}$$

$$\mathbf{R}_{\mathrm{H}}^{-1}\mathbf{R}_{\mathrm{K}} = \mathbf{R}_{\mathrm{T}} \qquad \text{......(56)}$$

were $\mathbf{Q}_{\mathrm{H}}$, $\mathbf{Q}_{\mathrm{K}}$ are symmetric (not necessarily diagonal) and $\mathbf{R}_{\mathrm{H}}$, $\mathbf{R}_{\mathrm{K}}$, $\mathbf{R}_{\mathrm{T}}$ are orthogonal matrices.

The unit normal $\mathbf{m}=m_i\hat{\mathbf{x}}_i$ to the plane in the parent from which the plane of coherence in the product is generated is assumed to be given. If $\mathbf{m}^*/|\mathbf{m}^*|$ is the unit normal to this plane of coherence then

$$\mathbf{m}^{*\prime}=\mathbf{m}'\mathbf{D}_{\mathrm{H}}^{-1}=\mathbf{m}'\mathbf{D}_{\mathrm{K}}^{-1}. \qquad \ldots\ldots(57)$$

Since $\mathbf{D}_{\mathrm{H}}$, $\mathbf{D}_{\mathrm{K}}$ produce coherent regions of the same structure differently oriented, det $(\mathbf{D}_{\mathrm{K}}\mathbf{D}_{\mathrm{H}}^{-1})=1$ and $\mathbf{D}_{\mathrm{K}}\mathbf{D}_{\mathrm{H}}^{-1}$ is thus a simple shear on the plane whose normal is $\mathbf{m}^*$. Let $\mathbf{l}^*$ be a vector along the direction of shear. The vector $\mathbf{l}$ from which $\mathbf{l}^*$ is generated is the required unit vector along the equivalent slip direction in the parent, and

$$\mathbf{l}^*=\mathbf{D}_{\mathrm{H}}\mathbf{l}=\mathbf{D}_{\mathrm{K}}\mathbf{l}. \qquad \ldots\ldots(58)$$

Two methods of procedure are now available according to the information assumed about the relative orientation of the regions H and K. In one type of analysis (e.g. Lieberman, Wechsler and Read 1955) it is assumed that $\mathbf{Q}_{\mathrm{H}}$ and $\mathbf{Q}_{\mathrm{K}}$ are known explicitly (from the correspondences for the H, K regions) and $\mathbf{R}_{\mathrm{T}}$ is calculated by considering the effect of $\mathbf{Q}_{\mathrm{H}}$, $\mathbf{Q}_{\mathrm{K}}$ on vectors which are to lie ultimately in the common boundary. Thus

$$\mathbf{D}_{\mathrm{H}}^{-1}\mathbf{D}_{\mathrm{K}}=\mathbf{Q}_{\mathrm{H}}^{-1}\mathbf{R}_{\mathrm{T}}\mathbf{Q}_{\mathrm{K}} \qquad \ldots\ldots(59)$$

may be calculated explicitly. However, the form of $\mathbf{D}_{\mathrm{H}}^{-1}\mathbf{D}_{\mathrm{K}}$ is also known in terms of $\mathbf{l}$, $\mathbf{m}$ and the twinning shear s. For

$$(\mathbf{D}_{\mathrm{K}}\mathbf{D}_{\mathrm{H}}^{-1})_{ij}-\delta_{ij}=s(l_i{}^*m_j{}^*)/|\mathbf{l}^*||\mathbf{m}^*| \qquad \ldots\ldots(60)$$

and hence

$$(\mathbf{D}_{\mathrm{H}}^{-1}\mathbf{D}_{\mathrm{K}})_{ij}-\delta_{ij}=s(l_im_j)/|\mathbf{l}^*||\mathbf{m}^*|. \qquad \ldots\ldots(61)$$

By comparison of (59) and (61), $A_i=sl_i/|\mathbf{l}^*||\mathbf{m}^*|$, and hence $l_i=A_i/|\mathbf{A}|$, and so s are easily found. An explicit expression for s can be obtained at once by noting that $\mathbf{D}_{\mathrm{K}}\mathbf{D}_{\mathrm{H}}^{-1}$ carries the vector $\mathbf{m}^*$ to $(\mathbf{D}_{\mathrm{K}}\mathbf{D}_{\mathrm{H}}^{-1})\mathbf{m}^*$. Thus

$$s=\tan\theta \qquad \ldots\ldots(62)$$

where $\cos\theta$ is the scalar product of $(\mathbf{D}_{\mathrm{K}}^{-1})'\mathbf{m}$ with $\mathbf{D}_{\mathrm{K}}\mathbf{D}_{\mathrm{H}}^{-1}(\mathbf{D}_{\mathrm{H}}^{-1})'\mathbf{m}$ divided by the product of the moduli of these vectors:

$$\cos\theta=\left(\frac{\mathbf{m}^1\mathbf{Q}_{\mathrm{H}}^{-2}\mathbf{m}}{\mathbf{m}^1\mathbf{Q}_{\mathrm{H}}^{-2}\mathbf{Q}_{\mathrm{K}}^{2}\mathbf{Q}_{\mathrm{H}}^{-2}\mathbf{m}}\right)^{1/2}. \qquad \ldots\ldots(63)$$

If the product is a stack of fine twins, and the fraction of material of type K is x, then x, s and the g of the present analysis are related by

$$sx=g. \qquad \ldots\ldots(64)$$

It is to be noted that, if $\mathbf{Q}_{\mathrm{H}}$, $\mathbf{Q}_{\mathrm{K}}$ are given, (57) and (58) imply the conditions

$$\mathbf{l}'(\mathbf{Q}_{\mathrm{K}}^{2}-\mathbf{Q}_{\mathrm{H}}^{2})\mathbf{l}=0 \qquad \ldots\ldots(65)$$

$$\mathbf{m}'(\mathbf{Q}_{\mathrm{K}}^{-2}-\mathbf{Q}_{\mathrm{H}}^{-2})=0\mathbf{m} \qquad \ldots\ldots(66)$$

restricting $\mathbf{m}$ and $\mathbf{l}$.

In a more general discussion Mackenzie and Bowles (1954) obtain an explicit expression for $\mathbf{D}_{\mathrm{H}}^{-1}\mathbf{D}_{\mathrm{K}}$ in terms of $\mathbf{Q}_{\mathrm{H}}$ and $\mathbf{m}$ only. The basis $\hat{\mathbf{x}}_i$ becomes in the region H the set $\mathbf{H}_i=(D_{\mathrm{H}})_{ji}\hat{\mathbf{x}}_j$; thus the bases P, P_f of the BM analysis may be identified with $\hat{\mathbf{x}}_i$, $\mathbf{H}_i$ respectively, and the BM quantities $\mathbf{M}$ and $\mathbf{h}$ with the present $\mathbf{Q}_{\mathrm{H}}$ and $\mathbf{m}$. This means that $\hat{\mathbf{x}}_i$ is chosen so that $\mathbf{Q}_{\mathrm{H}}$ is diagonal, as required in § 3. In twins of the first kind (an analogous development may be given for Type II twins) the structure in the K region is related to that in the H region by reflection in the plane with normal $\mathbf{m}^*$. Let $\mathbf{K}_i = (K|H)_i^j \mathbf{H}_j$ be the reflection of $\mathbf{H}_i$ in $\mathbf{m}^*$. Then $\mathbf{K}_i$ is generated from the same basis $\mathbf{y}_i$ in the parent, where $\mathbf{y}_i = (y|\hat{x})_i^j \hat{\mathbf{x}}_j$, and BM show that if $(y|\hat{x})_i^j$ represents a symmetry operation, it must be a reflection in the plane with normal $\mathbf{m}$. Thus the relative deformation of the two twins is given by

$$\mathbf{H}_i \rightarrow \hat{\mathbf{x}}_i = (\hat{x}|y)_i^j \mathbf{y}_j \rightarrow (\hat{x}|y)_i^j \mathbf{K}_j = (\hat{x}|y)_i^j (K|H)_j^k \mathbf{H}_k ;$$

the matrix $(\hat{\mathbf{x}}|\mathbf{y})(\mathbf{K}|\mathbf{H})$ is easily calculated in terms of $\mathbf{m}$ and $\mathbf{Q}_\mathrm{H}$. In fact,

$$(P_f \mathbf{S}^\mathrm{T} P_f)_{ij} = (\mathbf{D}_\mathrm{H}{}^{-1} \mathbf{D}_\mathrm{K})_{ij} = [(\hat{\mathbf{x}}|\mathbf{y})(\mathbf{K}|\mathbf{H})]_{ji} \qquad \ldots\ldots(67)$$

$$= \delta_{ij} + 2 v_i m_j \qquad \ldots\ldots(68)$$

where

$$v_i = (\mathbf{Q}_\mathrm{H}{}^{-2} \mathbf{m})_i / (\mathbf{m}' \mathbf{Q}_\mathrm{H}{}^{-2} \mathbf{m}) - m_i . \qquad \ldots\ldots(69)$$

The matrix $(P_f \mathbf{S}^\mathrm{T} P_f)$ given by (68) and (69) is the twinning shear in BM's equation (7.8). The direction of shear is thus parallel to the vector $v_i \mathbf{H}_i$ which is generated from the vector $v_i \hat{\mathbf{x}}_i$. In fact, this is true with v_i given by (69) even when $\hat{\mathbf{x}}_i$ is not chosen so that $\mathbf{Q}_\mathrm{H}$ is diagonal. BM also give an expression for the magnitude of the twinning shear. This is easily seen to be identical with (63) by using the relation

$$\mathbf{Q}_\mathrm{K}{}^2 = (P_f \mathbf{S}^\mathrm{T} P_f)' \mathbf{M}^2 (P_f \mathbf{S}^\mathrm{T} P_f) \qquad \ldots\ldots(70)$$

which follows from (67).

Finally, it may be remarked that although the above discussion is given for a twinned product, composite products not composed of twin-related structures are possible in principle and analogous developments may be given for them.

The application of the present theory to discuss the crystallography of particular transformations will be carried out in subsequent papers.

ACKNOWLEDGMENTS

One of us (R. B.) wishes to thank the University of Sheffield for the award of the C. G. Carlisle Scholarship. The other author (B. A. B.) is indebted to the Royal Society for the award of the Sorby Research Fellowship.

REFERENCES

BILBY, B. A., 1955, *Report of Bristol Conference on Defects in Crystalline Solids* (London : Physical Society), p. 124.

BILBY, B. A., BULLOUGH, R., and SMITH, E., 1955, *Proc. Roy. Soc.* A, **231,** 263.

BILBY, B. A., and CHRISTIAN, J. W., 1955, *Moncg. Ser. Inst. Metals*, No. 18, 121.

BOWLES, J. S., and MACKENZIE, J. K., 1954 a, *Acta Met.*, **2,** 129; 1954 b, *Ibid.*, **2,** 224.

FRANK, F. C., 1953, *Acta Met.*, **1,** 15.

LIEBERMAN, D. S., WECHSLER, M. S., and READ, T. A., 1955, *J. Appl. Phys.*, **26,** 473.

MACKENZIE, J. K., and BOWLES, J. S., 1954, *Acta Met.*, **2,** 138.

READ, W. T., 1953, *Dislocations in Crystals* (New York : McGraw-Hill).

WECHSLER, M. S., LIEBERMAN, D. S., and READ, T. A., 1953, *Trans. Amer. Inst. Min. Met. Engrs*, **197,** 1503.

Note added in proof :—One of us (R.B.) has recently remarked that (31) can be obtained without introducing g explicitly, although this quantity is, of course, required later. Put $k=1$ in (9), premultiply by **P**, place the term $\mathbf{Pp}^{(1)}$ on the left, and then multiply each side of the equation by its own transpose. Substitution for $\iota^{(1)}$ from (27) in the resulting equation then gives (31).

The Influence on Electric Strength of Mechanical Deformation caused by Application of Voltage to Potassium Chloride Crystals

BY R. COOPER AND A. A. WALLACE

Electrical Engineering Department, University of Manchester

MS. received 4th *May* 1956, *and in final form* 14th *August* 1956

Abstract. When electric fields of about $10^6\,\mathrm{v\,cm^{-1}}$ are applied to solid dielectrics the mechanical pressures developed between the electrodes are considerable. The manner in which this mechanical pressure may have influenced the measurements must be considered when interpreting experimental investigations of intrinsic electric strength. With many potassium chloride crystals it is shown that deformation of the crystal structure may be caused by application of the voltage before electrical breakdown takes place. The result of this is to increase appreciably the electric strengths relative to the values obtained from undeformed specimens. Information is given in the paper about the magnitude of the electric field in which structural deformation first takes place and about the subsequent behaviour of specimens in which structural defects have been introduced by the application of electric fields.

---

§ 1. INTRODUCTION

THE spread in values of electric strength obtained from apparently similar specimens of alkali halide crystals is very many times greater than is the value expected from the known measurement errors, which generally do not exceed about ± 10%. It is necessary to know the causes of the spread before any satisfactory interpretation can be given to investigations of intrinsic electric strength using these materials.

Calderwood, Cooper and Wallace (1953) have shown that mechanical deformation causes an increase in the average electric strength of potassium chloride and also that a suitable annealing treatment will reduce the average electric strength of deformed specimens. On the basis of this result they suggested that the wide spread in values of electric strength obtained from apparently similar specimens was due mainly to variations in the density of dislocations and other lattice defects which may be introduced during the preparation and handling of the specimens, and which may scatter conduction electrons and thus increase the electric strength. The ability of these materials to resist plastic deformation is not great.

More recently, Cooper and Wallace (1953) have demonstrated that there exists another possible cause of deformation. It is the mechanical pressure developed by the attraction between the electric charges appearing on the surfaces of the crystals when electric fields somewhat less than the breakdown field are applied. It is possible, therefore, that the electric strength of a specimen changes during the measurement and in a manner determined by its mechanical properties. The mechanical properties of apparently similar alkali halide crystals are known to vary appreciably (Boas and Schmidt 1950). Therefore, it is possible that the phenomenon of field induced strains provides an explanation of the wide spread observed in values of electric strength even when the strains produced during the preparation are removed by heat treatment and when, subsequently, great care is taken in the handling of the specimens. The purpose of the present note is to give information about the magnitude of the electric field in which mechanical deformation first commences and about the subsequent behaviour of specimens in which strains have been induced by the electric field. The experiments to be described were all performed with KCl, but there is no reason to expect qualitatively different behaviour from other alkali halide crystals.

§ 2. THE ELECTRIC STRESS REQUIRED TO PRODUCE INTERNAL STRAIN

The method of investigation used at room temperature was similar to that described by Cooper and Wallace (1953). The photo-elastic method was used to detect birefringence caused by the development of internal strains and the electric field was applied in the form of 1:50 microsecond impulses of successively increasing amplitude. When each impulse was applied the highly stressed region of the specimen was observed through a microscope and a sudden, permanent increase in the amount of light transmitted through the optical system indicated the onset of slip and the production within the crystal lattice of internal strain.

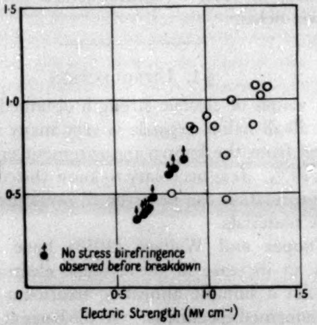


Figure 1. Relation between electric strength and (electric stress)^2 to induce stress. Birefringence in [100] specimens of KCl.

In figure 1 values of electric strength are plotted against the electric stress required to induce stress birefringence, i.e. $(E_{SB})_{min}$, in each group of twenty specimens. The electric field was applied in a [100] crystallographic direction. Each of this group of specimens was initially free from any strain that could be detected by inspection in a polarizing microscope, but despite this evidence of uniformity very considerable spread existed in the values of breakdown strength and also in the values of electric stress at which stress birefringence was first detected.

Eight of the specimens broke down before stress birefringence was detected and their electric strengths extended from 0·61 MV cm^{-1} to 0·85 MV cm^{-1}. The average value was 0·70 MV cm^{-1}. The information obtained from this experiment is summarized in the table along with similar information obtained using specimens in which the electric field was applied in the [110] direction, and in the [111] direction. Anisotropy of electric strength is clearly exhibited by KCl. This result is contrary to what is generally believed about the electric strengths of the alkali halides but the evidence of isotropy appears to consist only of sparse results obtained by Davisson and von Hippel (1940) using NaCl at room temperature.

Table. Electric Strength and Electric Stress to induce Stress Birefringence in KCl

(1)	(2)	(3)			(4)	(5)
		(*a*)	(*b*)	(*c*)		
[100]	20	0·96	1·34	0·61	0·68–1·07	0·61–0·85 (8 specimens)
[110]	19	1·26	1·50	0·90	0·75–1·22	none
[111]	24	1·19	1·53	0·82	0·80–1·06	0·82–0·94 (2 specimens)

(1) Type of specimen, (2) total number of specimens, (3) electric strength (MV cm^{-1}), (*a*) average, (*b*) maximum, (*c*) minimum, (4) range of values of electric stress at which stress birefringence was first observed (MV cm^{-1}), (5) range of electric stress (MV cm^{-1}) and number of specimens in which breakdown occurred before stress birefringence was observed.

§ 3. EFFECT OF TEMPERATURE

The apparatus used at room temperature enabled stress birefringence in the specimens to be detected the instant any deformation occurred. Consequently, there can be no doubt that the applied electric field was responsible for the observed effects and that the strains were not introduced accidentally by handling the specimens, but no satisfactory results were obtained using this apparatus at other temperatures. Nevertheless, an adequate assessment was obtained, by the following method, of the possible effects of temperature on the magnitude of the electric fields required to induce strains. The conventional electrode system and procedures for measuring electric strength were used (Calderwood, Cooper and Wallace 1953) but instead of applying a series of impulses until the specimen broke down only one impulse was applied to each specimen. The amplitude of each impulse was adjusted to yield the desired field strength and a subsequent inspection in a polarizing microscope revealed whether or not the impulse had strained the specimen. Five or six specimens were tested at each value of field strength. It was necessary to take precautions to ensure that no strains were produced by the thermal shock which occurred when a specimen was placed in an ambient at a temperature different from that of the room. Observations were made at 220°C and at −195°C. In the former case the specimen and its holder were heated together from room temperature and were allowed to cool again before removing the specimen. At −195°C the specimen was slowly lowered through the temperature gradient existing for seven or eight inches above the level of the liquid nitrogen used as ambient at −195°C. When the specimen was lowered too rapidly it caused vigorous boiling on entering the liquid and absence of this effect denoted a suitable lowering speed.

There was very little effect of temperature on the value of electric stress with which internal strains were first produced. Using (100) specimens it ranged from 0·6 MV cm^{-1} at 220°C to 0·7 MV cm^{-1} at −195°C and with (111) specimens the corresponding values were respectively 0·7 MV cm^{-1} and 0·80 MV cm^{-1}.

§ 4. WORK HARDENING

Since an electric discharge rather than a mechanical failure was always responsible for breakdown, it follows that some process within the crystal must set a limit to the increase in electric strength that can occur once the electric field has exceeded the value required to initiate deformation. This process seems to be connected with the phenomenon of work hardening, the presence of which was demonstrated in the following manner.

The change in light transmitted through the system of crossed polaroids and specimen was measured after each impulse was applied by attaching a photomultiplier tube to the eyepiece of the microscope. Measurements were made using fifteen (111) specimens of KCl. There was appreciable quantitative variation in the behaviour of the specimens but, in general, three patterns were observed and these were characterized by the curves (*a*), (*b*) and (*c*) shown in figure 2. In this diagram the total photomultiplier meter deflection was obtained by adding the deflection caused by a particular value of electric stress to the sum of all previous deflections. The curves qualitatively represent the mechanical stress–strain relationships for the crystals. Work hardening is clearly demonstrated in curves (*b*) and (*c*) and the highest values of electric strength were obtained from those specimens exhibiting the pronounced effect shown in curve (*c*). A group of specimens broke down very soon after the onset of stress birefringence and their behaviour is illustrated in curve (*a*). These specimens obviously suffered the least deformation and they had the lowest values of electric strength. A small number of specimens exhibited the rather anomalous behaviour illustrated in curve (*d*), where it is seen that a very large increase in light transmission followed each impulse right up to breakdown. As previously mentioned, very great spread existed amongst the values of electric stress required to initiate deformation. There was no correlation observed between this electric stress and the subsequent behaviour of annealed specimens, but if a specimen was purposely strained by mishandling, it behaved in the manner illustrated by curve (*e*). Pronounced work hardening occurred, the value of electric stress required to initiate deformation

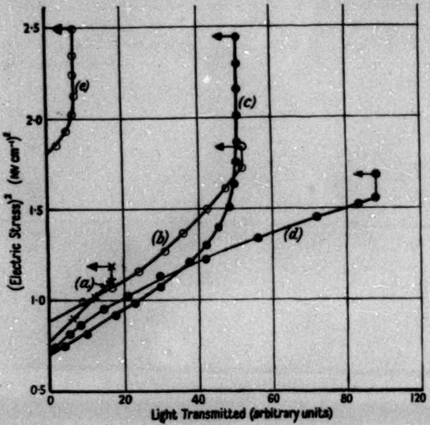


Figure 2. Work hardening of KCl during application of voltage. The arrows denote breakdown.

was high as was also the ultimate breakdown strength. At present the state of knowledge about plastic flow and the mechanical strength of crystals is in a state of flux and therefore any interpretation of the results described above is uncertain. According to current ideas the shear strength of a crystal is determined by the properties of dislocations which are considered to multiply and interact under the influence of a mechanical force and it is to this interaction that work hardening is attributed. From the above experiment it appears that a strained crystal, which contains more dislocations than does a crystal in which no strains can be detected, requires a greater force to cause plastic deformation than does a strain free crystal and it possesses a higher electric strength than does the latter type of crystal. From this it may be inferred that variations in the initial density of dislocations are likely to exist even in annealed crystals which appear in a polarizing microscope to be similar and free from internal strains because such crystals have been shown to vary appreciably in mechanical strength.

§ 5. CONCLUSIONS

The experiments leave no doubt about the occurrence in many cases of structural changes which produce an increase in electric strength and occur during the application of voltage to the specimen. One purpose of investigations of this kind is to provide data for comparison with theoretical predictions which are based on the assumption of a perfect crystal structure. No method exists of allowing quantitatively for the influence of strains on electric strength and therefore statements of maximum of average experimental values of electric strength are of little value for this purpose. The present experiments support the suggestion (Calderwood, Cooper and Wallace 1953) that the significant value of electric strength lies amongst the lower values and certainly is not the maximum value as was believed in the past.

Only the measurements made with the (100) specimens yielded values of electric strength that were not influenced by field induced strains. These values extended through a range of about 20% below the lowest value of electric stress that was found necessary to initiate deformation. This spread is greater than the known random errors and it may be asked why any field induced strains were ever observed with the (100) specimens? There are two probable reasons for this: First, the presence of inherent differences in the structure of the specimens and hence in their electric strength as discussed above and, second, the influence of time lags in breakdown. In many cases the time lag in breakdown may exceed 3 microseconds (Cooper and Grossart 1956) and this is enough to allow the voltage of a 1/50 microsecond impulse to decay sufficiently to make the application necessary of further impulses of greater amplitude, despite the fact that the field strength produced at the peak of the first impulse may have been sufficient to cause breakdown if it had been sustained for a longer period.

ACKNOWLEDGMENTS

The experiments described were carried out in the Electrical Engineering Laboratories of Liverpool University with the aid of a maintenance grant from the Department of Scientific and Industrial Research, for which we wish to express our thanks.

References

Boas, W., and Schmidt, E., 1950, *Plasticity in Crystals*, Chapter VII (London : F. A. Hughes).

Calderwood, J. H., Cooper, R., and Wallace, A. A., 1953, *Proc. Instn Elect. Engrs*, Part IIA, **100**, 105.

Cooper, R., and Grossart, D., 1956, *Proc. Phys. Soc.* B, **69**, 1351.

Cooper, R., and Wallace, A. A., 1953, *Proc. Phys. Soc.* B, **66**, 1113.

Davisson, J. W., and von Hippel, A., 1940, *Phys. Rev.*, **57**, 156.

Delayed Fracture of Sintered Alumina

By S. PEARSON

Royal Aircraft Establishment, Farnborough, Hants.

Communicated by L. G. Carpenter; *MS. received* 31*st May* 1956

Abstract. The strength of sintered alumina, when tested in the normal atmosphere, is shown to depend on the time for which the stress is applied. When stressed in bending under constant load, the average time to fracture increases from 1 sec to 10^6 sec when the stress is reduced by about 22%. This delayed fracture can be largely eliminated by heat treating and testing in a vacuum. The vacuum heat treatment does not change the strength of the material permanently since the delayed fracture effect reappears when the material is again exposed to the atmosphere. It is concluded that the effect is due to attack of the stressed material by constituents of the atmosphere.

---

§ 1. INTRODUCTION

It has been shown that the strength of mineral glass and of various vitreous porcelains tested in the normal atmosphere depends on the time for which the stress is applied (Holland and Turner 1940, Preston 1942, Baker and Preston 1946). Preston found that the average time required to fracture soda-lime–silica glass in bending under constant load increased from 10^{-2} to 10^4 seconds when the stress was reduced in the ratio 3 : 1. It has also been shown (Baker and Preston 1946, Gurney and Pearson 1949) that for mineral glass the effect is due to attack by constituents of the atmosphere (water vapour and carbon dioxide) and that the effect is negligible for glass tested *in vacuo* after prior vacuum heat treatment. Ways in which atmospheric attack can lead to delayed fracture have been discussed by Orowan (1944) and Gurney (1947).

It was considered by Gurney and Pearson (1949) that the delayed fracture in glass is due to attack of the highly stressed material at the end of Griffith cracks by constituents of the atmosphere derived either directly from the gaseous phase or from atmospheric constituents contained in the surface layers of the glass. The very high stresses developed in the material at the end of the cracks will alter the effective equilibrium with the atmosphere and so lead to change in strength. If this explanation is correct a delayed fracture should be obtained for all brittle materials which will react with the constituents of the atmosphere when stressed and the effect should not be confined to vitreous materials. It was considered likely that the stresses developed locally in brittle materials would be so great that practically all such materials would react with constituents of the atmosphere when stressed and, consequently, it was decided to investigate the delayed fracture in a brittle polycrystalline material which has very good resistance to chemical attack when stress free. Sintered alumina was chosen as a material which fulfilled these requirements.

§ 2. MATERIAL

The material used was sintered alumina for which the makers gave the following approximate composition: $Al_2O_3 \nless 98{\cdot}9\%$, $Fe_2O_3 \ngtr 0{\cdot}1\%$, $SiO_2 \ngtr 0{\cdot}1\%$, alkalies $\ngtr 0{\cdot}5\%$, Cu, Ti, Ca, Mg, trace or absent. It was received in the form of 200 rods, each 9 in. long by approximately ¼ in. diameter. The density of the rods was determined by weighing and measuring, and the mean density was found to be 3·66, with a standard deviation of 0·042. The density of single crystal alumina is 4·0 and the maximum density normally obtained for sintered material is 3·8–3·85. A sample from one of the rods was sectioned in the transverse and longitudinal directions and examined under the polarizing microscope. The grain size was reasonably uniform with a mean grain diameter of about $20\,\mu$. The grains showed some preferred crystallographic orientation, presumably due to the extrusion process used in the manufacture of the rods.

§ 3. EXPERIMENTS

Delayed fracture experiments were made in the normal laboratory atmosphere and afterwards experiments were made to see if the delayed fracture effect could be eliminated or reduced by vacuum heat treatment. The 200 test pieces were allocated to the different test conditions according to tables of random numbers.

3.1. *Apparatus*

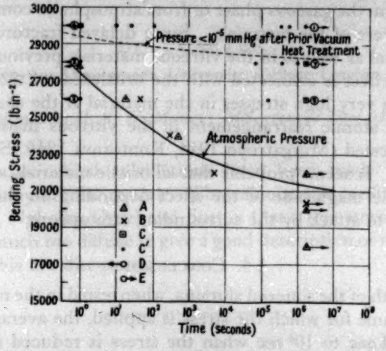
The apparatus used has been described previously (Gurney and Pearson 1949). Briefly the rods were tested in four point bending in flexible metal tubes. The tubes were open to atmosphere for the tests at normal atmospheric pressure, and were coupled up six at a time to a vacuum pumping system for the tests *in vacuo*. The vacuum pumps were kept running continuously during the experiments. The time to fracture was recorded automatically. Preliminary heating was done in electrically heated furnaces placed over the flexible metal tubes.

3.2. *Experiments in Normal Atmosphere with no prior Vacuum Heat Treatment*

Ten rods were loaded to failure in the normal atmosphere at a constant rate of increase of stress of about 30 000 lb in^{-2} min^{-1}. The mean failing stress in bending for the rods was 29 300 lb in^{-2} with a coefficient of variation of 14·1%. Delayed fracture tests were then made in which a series of constant loads was applied to the rods and the times to fracture were recorded. Twelve tests were made at each stress and the median times to fracture are represented by triangles in the figure.

3.3. *Experiments at a Pressure of less than* 10^{-5} *mm Hg after prior Vacuum Heat Treatment*

For this series of experiments the rods were first heated to 350°c ± 10°c for 48 hours in a vacuum of less than 10^{-5} mm Hg. They were then allowed to cool and the furnaces were removed. A series of constant stresses was then applied to the rods, and the times to fracture were recorded, the pressure being maintained at less than 10^{-5} mm Hg by continuous evacuation. Twelve tests were made at each stress and, as few delayed fractures occurred, all the results are shown in the figure. The corresponding median points are represented by squares.



Delayed Fracture of Sintered Alumina. A. Median points for experiments at atmospheric pressure with no prior vacuum heat treatment (12 tests at each stress). B. Median points for experiments at atmospheric pressure after prior vacuum heat treatment (6 tests at each stress). C. Median points and, C′, results of individual tests for experiments at pressure of $<10^{-5}$ mm Hg after prior vacuum heat treatment (12 tests at each stress). D. Number denotes number of specimens breaking during application of load. E. Number denotes number of specimens which did not break.

3.4. *Experiments in Normal Atmosphere after prior Vacuum Heat Treatment*

This series of experiments was made to see if the strength of the alumina was changed permanently by the vacuum heat treatment, or if the change in strength was reversible. Rods were heat treated at 350°c ± 10°c for 48 hours in a vacuum of less than 10^{-5} mm Hg as in the previous series of vacuum experiments. They were then allowed to cool, removed from the apparatus and allowed to stand in the normal laboratory atmosphere for at least 24 hours. A series of delayed fracture tests was then made on the rods in the normal atmosphere. Six rods were tested at each stress and the median times to fracture are represented by crosses in the figure. It will be seen that the crosses are scattered about the curve for the rods which had not been heat treated.

§ 4. Discussion of Results

The experimental results (see figure) show that the strength of the sintered alumina when tested in the normal atmosphere depends on the time for which the stress is applied, the average time to fracture increasing from 1 sec to 10^6 sec when the stress is reduced in the ratio 1·3 : 1. The comparable ratio for soda-lime–silica glass and for lead glass is about 2 : 1, for silica glass 1·7 : 1 and for vitreous porcelain 1·4 : 1. As was found for glass the delayed fracture effect can be largely eliminated by heat treating and testing in a vacuum. It is shown, however, that the vacuum heat treatment itself does not change the strength of the material since the delayed fracture curve for rods stored and tested in the normal atmosphere after vacuum heat treatment is not appreciably different from the curve for untreated material.

The results obtained in the experiments therefore support the hypothesis that the delayed fracture in brittle materials is due to attack of the highly stressed material at the end of Griffith cracks by constituents of the atmosphere derived either directly from the gaseous phase or from atmospheric constituents contained in the surface layers of the materials. Since delayed fracture occurs in a polycrystalline material as well as in the vitreous materials previously tested it seems evident that the effect is connected with the brittleness of the materials, i.e. the ability to develop very high stresses in the material at the sharp ends of cracks, rather than with atomic rearrangement in the vitreous materials under stress, as has been suggested (Murgatroyd 1944, Kontorova 1946, Saibel 1947, Taylor 1947, Cox 1948). It seems probable that all brittle materials will show a delayed fracture effect, the magnitude of the effect depending on the resistance of the stressed material to attack by the surrounding atmosphere.

§ 5. CONCLUSIONS

1. The strength of the sintered alumina, when tested in the normal atmosphere, depends on the time for which the stress is applied, the average time to fracture increasing from 1 sec to 10^6 sec when the stress is reduced in the ratio 1·3 : 1.

2. The delayed fracture effect can be largely eliminated by heat treating and testing in a vacuum, and it seems that it is due to attack of the stressed material by constituents of the atmosphere.

3. It is suggested that all brittle materials will show a delayed fracture effect, the magnitude of the effect depending on the resistance of the stressed material to attack by the surrounding atmosphere.

ACKNOWLEDGMENT

The author is much indebted to Professor C. Gurney, formerly of the Royal Aircraft Establishment, at whose instigation the work was undertaken.

REFERENCES

BAKER, T. C., and PRESTON, F. W., 1946, *J. Appl. Phys.*, **17**, 179.
COX, S. M., 1948, *J. Soc. Glass. Tech.*, **32**, 127.
GURNEY, C., 1947, *Proc. Phys. Soc.*, **59**, 169.
GURNEY, C., and PEARSON, S., 1949, *Proc. Phys. Soc.* B, **62**, 469.
HOLLAND, A. J., and TURNER, W. E. S., 1940, *J. Soc. Glass Tech.*, **24**, 46.
KONTOROVA, T. A., 1946, *Doklady Akad. Nauk, S.S.S.R.*, **54**, 1.
MURGATROYD, J. B., 1944, *J. Soc. Glass Tech.*, **28**, 406.
OROWAN, E., 1944, *Nature, Lond.*, **154**, 341.
PRESTON, F. W., 1942, *J. Appl. Phys.*, **13**, 623.
SAIBEL, E., 1947, *J. Chem. Phys.*, **15**, 760.
TAYLOR, N. W., 1947, *J. Appl. Phys.*, **18**, 943.

A Theoretical Study of Pentavalent Phosphorus

By C. CARTER
Physics Department, University College London

Communicated by H. S. W. Massey; *MS. received* 22*nd May* 1956

Abstract. Molecular orbital calculations have been made for the molecule PH_5 assuming a bipyramidal model. The calculations confirm the conclusion of Craig, Maccoll, Nyholm, Orgel and Sutton that a free atom 3d atomic orbital of phosphorus is much too diffuse to give a good description of molecular binding. The reasons for this are discussed.

§ 1. Introduction

The elements of the second short period differ from those of the first short period in that they can form compounds in which they have valence $8-V$, as well as those with their normal valence V. Thus, phosphorus forms compounds PF_3 and PF_5, whilst nitrogen forms only NF_3. A qualitative explanation of this behaviour can be given in terms of d-orbital hybridisation. The normal configuration of phosphorus $(3s)^2(3p)^3$ is trivalent, but by promotion of a 3s electron to 3d, a pentavalent configuration $(3s)(3p)^3(3d)$ is possible. The observed bipyramidal form of PF_5 can be explained by suitable hybridization of the five valence orbitals. However, nitrogen has a normal trivalent configuration $(2s)^2(2p)^3$ and promotion of the 2s electron to 3d is energetically more difficult.

In a recent study of chemical bonds involving d orbitals, Craig, Maccoll, Nyholm, Orgel and Sutton (1954) have attempted a semi-quantitative study of d-orbital hybridization in PCl_5 and SF_6, using atomic orbitals defined by Slater's rules and taking the overlap integral as a criterion of bond strength. In both cases, they found that the Slater 3d atomic orbital is much too diffuse to hybridize effectively with the 3s and 3p atomic orbitals. This result suggests either that d-orbital hydridization does not take place or else that it is a very bad approximation to use free atomic 3d orbitals in forming a molecular wave function.

Craig *et al.* suggest reasons why the d atomic orbital might be contracted by the peripheral atoms when a molecule is formed but, in the absence of a quantitative study, it is not possible to decide if these reasons are valid. Presumably, it would be possible to perform valence bond calculations in which the energy is minimized with respect to a variable parameter in the 3d atomic orbital, but they would be complicated by the need to include ionic configurations.

In this paper we attempt to throw light on the problem by indirect but more tractable calculations in which the molecule PH_5 is treated by an approximate molecular orbital method. PH_5 is apparently unstable but it is the simplest molecule to treat theoretically. Later, we hope to investigate the molecule PF_5 similarly and perhaps discover whether one should expect PF_5 to be more stable than PH_5.

§ 2. MOLECULAR ORBITAL WAVE FUNCTIONS FOR PH_5

The usual method of forming a molecular orbital wave function is to represent it as a linear combination of atomic orbital wave functions (LCAO). For molecules of high symmetry, however, it is possible to approximate the molecular orbital (MO) by the first term of a harmonic series about the centre of the molecule as origin. The radial wave functions may then be calculated by the S.C.F. method in the same way as for atoms. Buckingham, Massey and Tibbs (1941) showed that this method predicts some properties of methane quite satisfactorily. Carter (1956) has investigated the method in more detail and discussed the effect of including higher terms in the harmonic series for the MO. The single term approximation appears to give bond lengths quite accurately, but it fails to give quantitative binding energies.

The exact MO electron configuration of the ground state of PH_5 is

$$(1A_1')^2(2A_1')^2(2A_2'')^2(2E')^4(3A_1')^2(3A_2'')^2(3E')^4(4A_1')^2.$$

In the single harmonic term approximation this becomes (Carter 1955)

$$(1s)^2(2s)^2(2p)^6(3s)^2(3p)^6(4A_1')^2.$$

The correspondence is unambiguous except in the case of the $4A_1'$ orbital in which the $3d_{z^2}$ and 4s terms are likely to be of about the same importance. Best results would be obtained by taking a linear combination of d and s terms, but this course was not followed since it would complicate the calculations considerably and would destroy the simple relation between molecular orbitals and atomic orbitals.

Instead, we first solved the Hartree S.C.F. equations approximately for PH_5^{2+} with an argon-like electron configuration, taking the P–H bond length $r_0 = 2{\cdot}70$ A.U. The Hartree equations were then solved for a 3d and a 4s electron in the unperturbed field of PH_5^{2+}. These equations are

$$\left.\begin{aligned} &\frac{d^2P(3d)}{dr^2} + \left[10V_0 + \frac{2}{7}V_2 + \frac{25}{14}V_4 + \frac{2Z(r)}{r} - \frac{6}{r^2} - \epsilon_{3d}\right]P(3d)=0\\ &\frac{d^2P(4s)}{dr^2} + \left[10V_0 + \frac{2Z(r)}{r} - \epsilon_{4s}\right]P(4s)=0 \end{aligned}\right\} \quad \ldots\ldots(1)$$

where $Z(r)/r$ is the field due to the P nucleus and the electrons in PH_5^{2+} and

$$\begin{aligned} V_k(r) &= r^k/r_0^{k+1}, \quad \text{for } r \leqslant r_0,\\ &= r_0^k/r^{k+1}, \quad \text{for } r \geqslant r_0. \end{aligned} \quad \ldots\ldots(2)$$

The V_k terms represent the effective potential field due to the five H nuclei arranged bipyramidally (figure 1). The terms V_2 and V_4 in the 3d equation arise from the angular dependence of the 3d wave function. $\frac{1}{2}\epsilon$ is approximately equal to the binding energy of the outer electron in PH_5^+. The 3d equation was solved both with and without the non-central potential terms V_2 and V_4. The values of ϵ obtained were

$$\begin{aligned} \epsilon_{3d} \text{ (omitting } V_2 \text{ and } V_4) &= 0{\cdot}50\\ \epsilon_{4s} \qquad\qquad\qquad\qquad &= 0{\cdot}55\\ \epsilon_{3d} \text{ (including } V_2 \text{ and } V_4) &= 0{\cdot}62. \end{aligned}$$

We conclude therefore that, in contrast to Ca^+, a 3d electron in PH_5^+ is more tightly bound than a 4s electron, the tighter binding being largely due to the non-central interaction of the electron with the five protons.

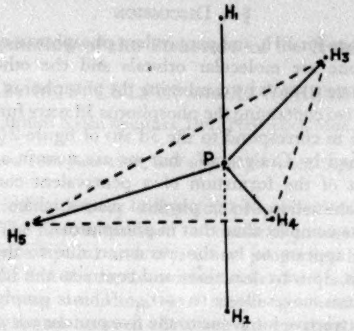


Figure 1. The bipyramidal model assumed for PH_5. The polar bond lengths $P–H_1$, $P–H_2$, have been taken equal to the equatorial bond lengths $P–H_3$, $P–H_4$, $P–H_5$.

With the above calculations on PH_5^+ as a guide, the $4A_1'$ orbitals in PH_5 were taken as $3d_{z^2}$ since, as in the atomic case, the relative binding of a 3d electron over a 4s is likely to increase with increasing number of electrons. This choice is further justified in that the error caused by omitting the s component is partially cancelled by the omission of the d component in the $3A_1'$ orbital.

The calculated S.C.F. wave functions for PH_5 are drawn in figure 2. For comparison, wave functions for pentavalent phosphorus have been calculated and are also shown in figure 2. It can be seen that the 3d atomic orbital is very extended, as predicted by Slater's rules.

In view of the approximate nature of the calculations, only rough self-consistency was aimed at but, on the scale of the figure, the self-consistent wave functions would differ insignificantly from those drawn.

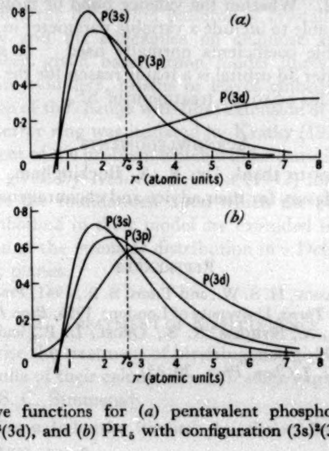


Figure 2. Radial wave functions for (*a*) pentavalent phosphorus with configuration $(3s)(3p)^3(3d)$, and (*b*) PH_5 with configuration $(3s)^2(3p)^6(3d)^2$.

§ 3. DISCUSSION

The wave functions for PH_5 and pentavalent phosphorus cannot be compared directly since the one are molecular orbitals and the other atomic orbitals. However, if we form LCAO MO's by combining the phosphorus AO's with hydrogen AO's, we see that any MO containing the phosphorus 3d wave function of figure 2 (*a*) would be too diffuse to correspond to the 3d MO of figure 2 (*b*). This confirms the conclusion reached by Craig *et al.*, but we are now in a position to offer a physical explanation of the formation of a pentavalent compound. All that is required is an examination of the physical causes which make the 3d wave function in PH_5 more compact than that in phosphorus.

The chief cause appears to be the attraction due to the hydrogen atoms, which expands the 3s, 3p wave functions and contracts the 3d function so that in each case the maximum moves closer to $r = r_0$. This is simply an electronegative effect, in which the attractive force due to the five protons outweighs the repulsion due to the five extra electrons, and might be expected to be larger when the hydrogen atoms are replaced by more electronegative ligands. This effect is accentuated by the electron pairing. The single 3d electron in phosphorus is strongly screened from the nuclear charge by the inner electrons, but one 3d electron in PH_5 is only slightly screened by the other one. This means that a 3d electron is more strongly attracted towards the centre in PH_5 than in phosphorus, and is accordingly more compact. The effect is cumulative since a contraction of the 3d electron cloud reduces the screening caused by the inner electrons and hence causes a further contraction.

§ 4. CONCLUSION

It seems probable that d hybridization does occur in both pentavalent phosphorus and hexavalent sulphur. The 3d atomic orbital is, however, much contracted during molecular formation and when making calculations with molecular wave functions involving atomic orbitals it is not satisfactory to use a free atom 3d orbital. Whether the valence bond or molecular orbital method is employed it is desirable to include a variable parameter in the 3d wave function as well as the variable coefficients normally used. It seems likely that the inadequacy of the Slater 3d orbital is a major reason for the poor results obtained by Duncan (1952) in his LCAO MO treatment of SF_6.

ACKNOWLEDGMENTS

The author wishes to thank Dr. R. A. Buckingham, Dr. A. Maccoll and Professor H. S. W. Massey for their advice and encouragement during the course of this work.

REFERENCES

BUCKINGHAM, R. A., MASSEY, H. S. W., and TIBBS, S. R., 1941, *Proc. Roy. Soc.* A, **178**. 119.

CARTER, C., 1955, *Ph.D. Thesis*, University of London; 1956, *Proc. Roy. Soc.* A, **235**, 321.

CRAIG, D. P., MACCOLL, A., NYHOLM, R. S., ORGEL, L. E., and SUTTON, L. E., 1954, *J. Chem. Soc.*, 332.

DUNCAN, A. B. F., 1952, *J. Chem. Phys.*, **20**, 951.

The Orientation of the Structure of 6·6 Nylon Fibres

BY M. F. CULPIN AND K. W. KEMP
British Nylon Spinners, Ltd., Pontypool, Mon.

MS. received 28th June 1956, and in final form 6th September 1956

Abstract. The optical and x-ray diffraction properties of a model fibre have been considered. The fibre is assumed to consist of an assemblage of structural units to which simple optical and crystalline properties have been assigned. It is further assumed that the random orientation of units in the undrawn fibre is modified by drawing according to a simple geometrical pattern. Good agreement has been found between the observed variation of the birefringence with drawing, and the variation calculated from the model, and between some of the observed changes with drawing of the x-ray diffraction pattern, and the corresponding changes shown by the model.

§ 1. INTRODUCTION

THE production of a fibre of 6·6 nylon consists essentially of two processes. Firstly, the chip polymer is melted and extruded through a small hole to give the material the required form. After solidification, the fibre is extended or drawn far beyond its elastic limit to several times its spun length in order to give it the stiffness required in a textile fibre. We are concerned in this paper with what happens to the fine structure of the material during the second process.

X-ray examination shows 6·6 nylon to be a largely crystalline material, of which part is amorphous. Systematic study of the effect of large extensions upon the birefringence and x-ray diffraction patterns of crystalline fibres began with the work of Kratky and his colleagues on cellulose and similar materials. Calculations have been given based upon model fibres, assumed to consist variously of crystallite cylinders, prisms or leaves embedded in an amorphous matrix. A calculation of the change with fibre extension of the intensity distribution in a Debye–Scherrer ring was initiated by Kratky (1933), but his work falls short of a full treatment of the problem which will be considered here.

The expressions given by Kuhn and Grün (1942) for the birefringence of a crystalline fibre are the foundations upon which the present work is based. The assumptions embodied in their model are extended in the present work to permit the calculation of the intensity distribution in a Debye–Scherrer ring due to a set of paratropic planes.

§ 2. TREATMENT OF BIREFRINGENCE

We shall summarise the treatment of birefringence given by Kuhn and Grün, and compare the results of their calculation with some experimental work undertaken for us by Mr. S. C. Simmens†.

---

† Now at British Cotton Industries Research Association, Didsbury, Manchester.

Following Kuhn and Grün, we assume the model fibre to be cylindrical and to consist of an assemblage of structural units with optical polarizability δ_1 along the axis of the unit and δ_2 in a plane perpendicular to this direction. The units are assumed to be randomly oriented in three-dimensional space when the fibre is in the undrawn state. The number dG of units per unit volume with axes inclined at an angle β to the fibre axis is equal to $\frac{1}{2}G\sin\beta d\beta$, where G is the number density of the units.

We choose rectangular cartesian axes x, y, z, within the fibre, with the z axis parallel to the fibre axis. We assume that the direction of the axis of a structural unit which is defined by the line joining the points (x_1, y_1, z_1) and (x_2, y_2, z_2) before drawing will change to that of the line joining the two points $(x_1/\sqrt{\lambda}, y_1/\sqrt{\lambda}, \lambda z_1)$ and $(x_2/\sqrt{\lambda}, y_2/\sqrt{\lambda}, \lambda z_2)$ after drawing. To understand the relevance of the transformed coordinates, we note that, under the assumption of no change in volume, a piece of the fibre of length L and radius r before drawing, has length λL and radius $r/\sqrt{\lambda}$ after drawing to a ratio of λ. It is easily shown that the number dG of units inclined at an angle β to the fibre axis after the extension is given by

$$dG = \frac{G\lambda^3 \sin\beta d\beta}{[1+(\lambda^3-1)\sin^2\beta]^{3/2}}. \qquad \ldots\ldots(1)$$

Further, it may be shown that at this draw ratio the birefringence of the system is given by

$$\Delta n(\lambda) = \frac{2\pi G}{9n(\lambda)}[n^2(\lambda)+2]^2(\delta_1-\delta_2)f(\lambda) \qquad \ldots\ldots(2)$$

where

$$f(\lambda) = \frac{2\lambda^3+1}{2(\lambda^3-1)} - \frac{3\lambda^3 \tan^{-1}\sqrt{(\lambda^3-1)}}{2(\lambda^3-1)^{3/2}}$$

and $n(\lambda)$ is the mean refractive index at an extension of λ. The birefringence is assumed to be small compared with $n(\lambda)$. Since $f(\lambda)\to 1$ as $\lambda\to\infty$, we may write

$$\Delta n(\lambda) = \Delta n(\infty) f(\lambda) \qquad \ldots\ldots(3)$$

where $\Delta n(\infty)$, the maximum attainable birefringence, contains the factors G, $\delta_1-\delta_2$ and $[n^2(\lambda)+2]^2$. The first two factors cannot at present be measured independently. The last arises because we have used the Lorentz–Lorenz equation for obtaining the refractive index from the optical polarizability. There is justification (Brown 1940) for using the simpler relation $n^2-1=4\pi G\sigma$ where n is the refractive index, σ the optical polarizability and G the number density of polarizable units. In this case we could obtain for $\Delta n(\lambda)$ the expression

$$\Delta n(\lambda) = \frac{2\pi G}{n(\lambda)}(\delta_1-\delta_2)f(\lambda). \qquad \ldots\ldots(4)$$

For 6·6 nylon, however, both $[n^2(\lambda)+2]^2/n(\lambda)$ and $1/n(\lambda)$ may be regarded as constants in this work.

We proceed to compare equation (3) with some experimental results which were obtained by the following technique. Fibres were drawn to different ratios under a variety of conditions of speed and temperature. The samples were allowed to relax for twenty-four hours after drawing. After this interval it was found that the birefringence remained sufficiently constant in time for measurements to be made, and that changes that occurred thereafter due to further relaxation could be entirely accounted for in terms of diameter change

After the period of relaxation the samples were mounted in liquid paraffin for examination under the microscope and their birefringence measured in sodium light with the use of a Berek compensator. To do this the diameter and retardation were measured at a number of points on each sample. On the basis of previously observed cross sections of similar samples, the assumption of circularity was made.

In figure 1 we plot birefringence against draw ratio for one particular set of drawing conditions, together with a curve obtained with the use of equation (3), taking for $\Delta n(\infty)$ the value 0·073. This value was chosen to make the straightest parts of the corresponding curves of birefringence against diameter coincide as nearly as possible, as in figure 2. The curves for other drawing conditions showed similar agreement when small adjustments were made in the value of $\Delta n(\infty)$. Insufficient work has been done to establish any systematic variation of $\Delta n(\infty)$ with drawing conditions. It seems, however, that equation (3) may be applied equally well to drawing which has taken place evenly over the whole length and to drawing that has taken place at a neck (Marshall and Thompson 1954).

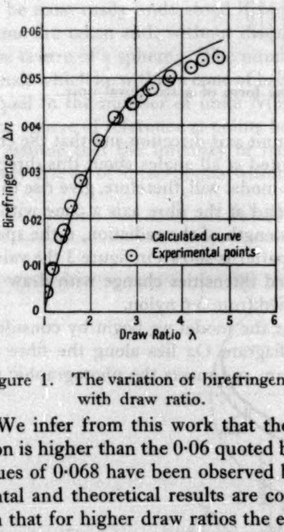


Figure 1. The variation of birefringence with draw ratio.

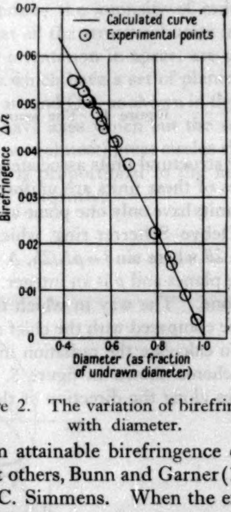


Figure 2. The variation of birefringence with diameter.

We infer from this work that the maximum attainable birefringence of 6·6 nylon is higher than the 0·06 quoted by, amongst others, Bunn and Garner (1947). Values of 0·068 have been observed by Mr. S. C. Simmens. When the experimental and theoretical results are compared in the manner described it will be seen that for higher draw ratios the experimental curves tend to fall below those obtained for the model fibre. One explanation could be that the manner of orientation may be in some way less 'perfect' than that assumed.

§ 3. X-RAY DIFFRACTION PATTERNS

It has been established by Bunn and Garner (1947) that the unit cell of 6·6 nylon is triclinic, that the *c* axis becomes oriented towards the fibre axis when the fibre is drawn, and that the most prominent features of the X-ray diffraction pattern by transmission are the ring due to the (100) plane and that due to both the (110) and (010) planes. The rings due to the last two unfortunately coincide. A pattern typical of the undrawn fibre is shown in figure 3 (i), whilst figures 3 (ii) and (iii) show the progressive effect of drawing upon this pattern (Plate).

We now invest our structural unit with crystalline properties and proceed to obtain an expression for the intensity of radiation at a point P on a Debye–Scherrer ring of the model fibre at any draw ratio λ. We assume that each unit contains a set of equidistant crystal planes lying parallel to the unit axis (see figure 4) and that the unit axis is parallel to the c axis. It is supposed that there are many structural units associated with each unit axis direction and that the crystal planes of these units are uniformly distributed at all angles about this direction. The units have only one plane spacing: our model will, therefore, give rise to only one Debye–Scherrer ring, which will subtend at the fibre axis a cone with half angle 2θ where $\sin\theta = p\Lambda/2h$, Λ is the wavelength of the radiation, h the spacing of the planes and p is an integer. For the patterns shown in figure 3 the value of p is one. The way in which the calculated intensities change with draw ratio will be compared with the chief rings obtained from 6·6 nylon.

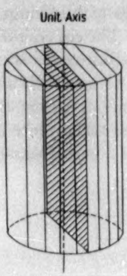


Figure 4. The assumed crystalline form of a structural unit.

To calculate the radiation intensities for the model we begin by considering the schema shown in figure 5. In this diagram Oz lies along the fibre axis. Ox lies along the direction of the x-ray beam and meets the photographic plate

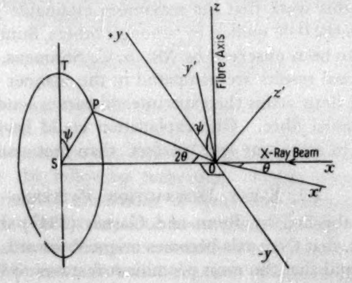


Figure 5. Diagram to illustrate the formation of the x-ray difiraction pattern.

which is perpendicular to it at S. Oy is perpendicular to both Ox and Oz. ST is parallel to Oz and P is the point at which we wish to calculate the intensity of radiation. $\mathrm{T\hat{S}P} = \psi$.

In the plane yOz, draw Oy' such that $y\hat{\mathrm{O}}y' = \psi$, and in the plane xOP draw Ox' such that $x\hat{\mathrm{O}}x' = \theta$. It can be shown that $x'\hat{\mathrm{O}}y' = 90°$. Draw O$z'$ perpendicular to the plane x'Oy'. This construction is shown in figure 5.

All sets of crystal planes which contribute to the diffracted intensity at P have Oz' as a normal. We therefore calculate the number density of structural units which have a set of planes perpendicular to Oz'. If we consider a unit whose axis is inclined at an angle β to Oz and for which the projection of this line on the xOy plane makes an angle ϕ with Ox, then the number density dG of such units in the undrawn state is given by $dG = G\sin\beta\, d\beta\, d\phi/4\pi$ where $0 \leqslant \beta \leqslant \frac{1}{2}\pi$, and $0 \leqslant \phi \leqslant 2\pi$. When the fibre is drawn to a draw ratio of λ we obtain, in accord with equation (1),

$$dG = \frac{G\lambda^3 \sin\beta\, d\beta\, d\phi}{4\pi[1+(\lambda^3-1)\sin^2\beta]^{3/2}} \equiv F(\lambda, \beta)\, d\beta\, d\phi. \qquad \ldots\ldots(5)$$

The calculation which we have done to estimate the intensity of radiation may be most easily understood if we imagine that all the structural units in unit volume are taken and, without disturbing their orientation in space, are placed at the centre of a sphere. The number of units which have a set of planes with normals which lie within a cone of half angle $\frac{1}{2}d\beta$ around Oz', as shown in figure 6, is equal to the number of units $N(\lambda, \psi)$ which have axes which cut the surface of the sphere of reference at points between the two (almost) great circles shown. We assume that the intensity of radiation at P is proportional to the number density $I(\lambda, \psi)$ of these points within the zone. Evidently,

$$I(\lambda, \psi) = \frac{N(\lambda, \psi)}{2\pi\, d\beta}. \qquad \ldots\ldots(6)$$

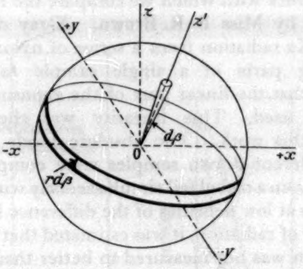


Figure 6. The sphere of reference with the distribution of axes of the units which contribute to the intensity at P.

An expression for $I(\lambda, \psi)$ may be obtained most easily by working with the coordinate axes x', y', z', of figure 5. The coordinates (β, ϕ) of a unit referred to the original set of axes and the corresponding coordinates (β', ϕ') referred to the new set are related by the equations

$$\left.\begin{aligned} \cos\beta &= \cos\theta\cos\psi\cos\beta' + \sin\psi\sin\beta'\sin\phi' - \sin\theta\cos\psi\sin\beta'\cos\phi'. \\ \cos\phi\sin\beta &= \cos\theta\sin\beta'\cos\phi' + \sin\theta\cos\beta'. \end{aligned}\right\}\ldots(7)$$

The Jacobian of (β, ϕ) with respect to (β', ϕ') is, in the usual notation, given by

$$\frac{\partial(\beta, \phi)}{\partial(\beta', \phi')} = \frac{\sin \beta'}{\sin^2 \beta \sin \phi} [\sin \psi \sin \theta \sin \beta' \cos \phi' + \cos \psi \sin \beta' \sin \phi' - \sin \psi \cos \theta \cos \beta'] \qquad \ldots\ldots(8)$$

To obtain the number $N(\lambda, \psi)$ of units having a set of planes with Oz' as a normal we put $\beta' = \frac{1}{2}\pi$, and integrate over all values of ϕ' from 0 to 2π. We have already $dG = F(\lambda, \beta)\, d\beta\, d\phi$ so that in the new coordinate system

$$dG = F(\lambda, \beta') \frac{\partial(\beta, \phi)}{\partial(\beta', \phi')} d\beta' d\phi' \qquad \ldots\ldots(9)$$

whence

$$N(\lambda, \psi) = d\beta' \int_0^{2\pi} F(\lambda, \beta') \frac{\partial(\beta, \phi)}{\partial(\beta', \phi')} d\phi' \quad [\beta' = \pi/2]$$

or

$$N(\lambda, \psi) = \frac{2 d\beta' \lambda^3}{\pi} \int_{-\pi/4}^{+\pi/4} \frac{d\alpha}{\{1 + \frac{1}{2}(\lambda^3 - 1)[(1+m) + (1-m)\sin 2\alpha]\}^{3/2}} \qquad \ldots(10)$$

where

$$m = \cos^2 \theta \cos^2 \psi, \quad \alpha = 2\phi' + \xi,$$
$$\xi = \arctan \{[1 - \cos^2 \psi (1 + \sin^2 \theta)]/2 \sin \psi \cos \psi \sin \theta\}.$$

This integral may be evaluated with the use of Jacobian elliptic functions, and, together with equation (6), gives

$$I(\lambda, \psi) = \frac{G\lambda^3 E[\sqrt{[(2b/(b+1)]}, \frac{1}{2}\pi]}{\pi^2 (1-b) \sqrt{[a^3(1+b)]}} \qquad \ldots\ldots(11)$$

where

$$a = 1 + \tfrac{1}{2}(\lambda^3 - 1)(1+m), \quad b = \tfrac{1}{2} a^{-1}(\lambda^3 - 1)(1-m),$$
$$E(k, \Phi) = \int_0^{\Phi} \sqrt{(1 - k^2 \sin^2 \Phi)}\, d\Phi.$$

The experimental work with which we compare the results of this calculation was undertaken for us by Miss B. R. Brown. X-ray diffraction patterns were obtained with copper Kα radiation from a series of nylon fibres which had been prepared by drawing parts of a single sample to different draw ratios. Exposures were such that the linear part of the exposure–density characteristic of the emulsion was used. This linearity was checked and found to be sufficiently exact for this work. The absolute levels of density of the rings obtained from the different drawn samples were compared by taking a back-reflection photograph from a nickel foil simultaneously with each fibre photograph. Owing to the smallness at low densities of the difference between the background and diffracted intensity of radiation, it was estimated that the absolute level of any one density distribution was not measured to better than 10%.

In order to compare the theoretical results embodied in equation (11) with the experimental results, we assume that $I(\lambda, \psi)$ is proportional to the intensity of radiation and that that is proportional to the density of silver deposit on the photographic plate. Since the constants of proportionality are not known, we have the choice of two procedures. Firstly, we may choose an arbitrary value of G and a scaling factor which will ensure coincidence of the calculated and observed curves for $\lambda = 1$ for each Debye–Scherrer ring. The density scale is then fixed for the calculated and observed curves for all other values of λ. Thus we obtain the curves shown in figures 7 and 8. Secondly, since the absolute level of the experimental curves is not accurately known, we may again select an arbitrary value of G, but then choose a different scaling factor for each value of λ to ensure coincidence of the observed and calculated curves at $\psi = 90°$. This has been done in figures 9 and 10.

It may be seen from figure 7 that the observed curves for the (110) plus (010) ring lie above the calculated curves. It appears from figures 8 and 10 that the (100) plane as incorporated in our structural unit does not follow the model

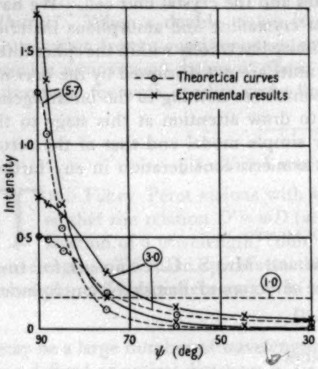


Figure 7. Experimental and calculated diffracted intensities for the (110) and (010) planes with the same scaling factor for all values of the draw ratio. Circled numbers give the draw ratio.

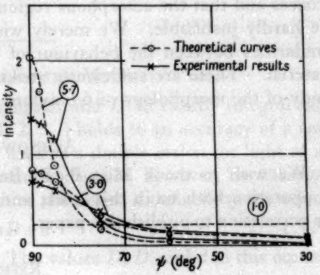


Figure 8. As in figure 7, but for the (100) plane.

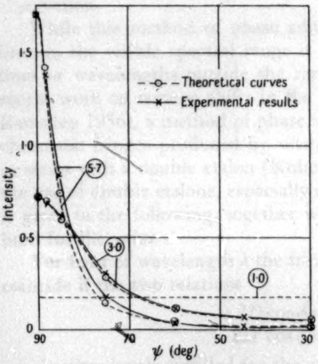


Figure 9. As in figure 7, but with the use of a different scaling factor for each value of draw ratio.

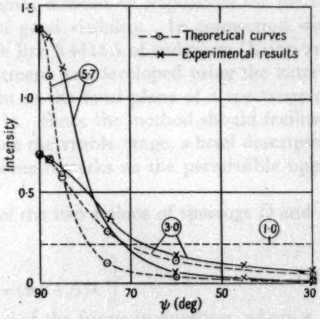


Figure 10. As in figure 9, but for the (100) plane.

deformation as closely as the (100) and (010) planes, for which the observed and calculated curves show good agreement.

§ 4. Discussion

The structure of 6·6 nylon fibres as seen under the microscope and the effect of drawing upon this structure are illustrated by Langkammerer and Catlin (1948). The structural units discussed in this paper may be considered to be intermediate in size between the spherulites and the crystal unit cell. We have not considered the relative proportions of crystalline and amorphous material, nor the question whether or not the crystallinity lies wholly within the spherulites. Again, the assumption that our structural units are not disrupted by the drawing process and that the amorphous regions contribute nothing to the birefringence are hardly justifiable. We merely wish to draw attention at this stage to the similarities between the behaviour of our simple model and that of the actual material. These are sufficiently striking to merit consideration in any further study of the morphology of 6·6 nylon.

Acknowledgments

We wish to thank Miss B. R. Brown and Mr. S. C. Simmens for their co operation with us in this work, and the directors of British Nylon Spinners for permission to publish this paper.

References

Brown, G. B., 1940, *Proc. Phys. Soc.*, **52,** 577.
Bunn, C. W., and Garner, E. V., 1947, *Proc. Roy. Soc.* A, **189,** 39.
Kratky, O., 1933, *Kolloid Z.*, **64,** 213.
Kuhn, W., and Grün, F., 1942, *Kolloid Z.*, **101,** 248.
Langkammerer, C. M., and Catlin, W. E., 1948, *J. Polymer Sci.*, **3,** 305.
Marshall, I., and Thompson, A. B., 1954, *J. Appl. Chem.*, **4,** 145.

The Phase Adjustment of the Double Etalon

By H. G. KUHN and S. A. RAMSDEN
Clarendon Laboratory, University of Oxford

MS. received 5th July 1956

Abstract. When a double etalon is used in conjunction with a spectrograph, illumination from a continuous source produces heterochromatic fringes; these provide a convenient means of phase adjustment. The method offers special advantages for weak or ultra-violet lines.

---

§ 1. The Principle of the Method

Two Fabry–Perot etalons with spacings D and D' in exactly integral ratio, so that the relation $D' = mD$ ($m = 1, 2, \ldots$) holds to an accuracy of a small fraction of a wavelength, could be used as a double etalon for light of any wavelength. To make spacers to this accuracy would, however, be extremely laborious and also unnecessary. In practice the defect d defined by the relation

$$D' = mD + d \qquad \ldots\ldots(1)$$

may be a large number of wavelengths. The values D, D' and d in this equation are defined as optical distances in air.

For any given wavelength of the light the fringe systems of the two etalons will then, in general, be out of step, and a 'phase adjustment' is required, consisting in a change of the path difference of one of the etalons by a fraction of a wavelength. This is most conveniently achieved by a slight change of air pressure (Jackson and Kuhn 1938); the values of the air pressure for which the intensity passes through maxima and minima may be established by visual observation.

While this method of phase adjustment has proved satisfactory for strong lines in the visible spectral range it becomes difficult or impossible for weaker lines or wavelengths outside the range of good visibility. In connection with recent work on isotope shifts in the spark line λ4416 Å of cadmium (Kuhn and Ramsden 1956), a method of phase adjustment was developed using the heterochromatic fringes produced by white light in the focal plane of a spectrograph in series with a double etalon (Kuhn 1951). Since the method should facilitate the use of double etalons, especially outside the visible range, a brief description is given in the following, together with some remarks on the permissible upper limit for the value d.

For light of wavelength λ the fringes of the two etalons of spacings D and D' coincide if the two relations

$$\left.\begin{aligned} 2D\cos\phi &= n\lambda \\ 2D'\cos\phi &= n'\lambda = (mn+\Delta)\lambda \end{aligned}\right\} \qquad \ldots\ldots(2)$$

are simultaneously fulfilled for the angle ϕ of the fringe in question, where n, n' and Δ are integral numbers. From (1) and (2) follows the relation

$$2d\cos\phi = \Delta\times\lambda \qquad \ldots\ldots(3)$$

which shows that the order difference Δ for the two etalons is the same as the order of an 'equivalent' single etalon of spacing d. Phase coincidence of the double etalon occurs if Δ has an integral value. This can be strictly achieved for only one value of ϕ, but if d is small enough a sufficiently close coincidence can be obtained for the whole range of ϕ covered by the fringe system. For this to be so, Δ should vary by only a small fraction of an order, perhaps by 1/20, depending on the reflectivity of the suppressor etalon. From (2) and (3) follows

$$(\Delta_0 - \Delta)/(n_0 - n) = d/D \qquad \ldots\ldots(4)$$

where the suffix o refers to the value of the order for the centre of the fringe pattern. Since $n_0 - n$ is normally about 1 to 3, d will usually have to be restricted to values below about $D/50$. It is generally easy to make etalon spacers to an accuracy of this order.

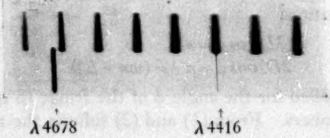
With a value of d of this order of magnitude, white light passing through the two etalons in series with a spectrograph will produce a sequence of well-resolved heterochromatic fringes, of the same appearance as those observed with a film of thickness d, bounded by partially reflecting surfaces. The maxima mark the wavelengths for which the etalons are in phase. The width of the peaks, relative to their spacing, depends mainly on the reflectivity of the plates of the suppressor etalon, and their spacing in wavelengths $\delta\lambda$, or in wave numbers $\delta\nu$, is given approximately by $|\delta\lambda/\lambda| = |\delta\nu/\nu| = -\lambda/2|d$.

§ 2. THE EXPERIMENTAL PROCEDURE

The slit of the spectrograph is illuminated simultaneously with white light and with light from the source emitting the spectral line to be investigated. The air pressure is then changed until one of the heterochromatic fringes coincides with the line. The phase adjustment is thus reduced from a setting for maximum or minimum intensity to the setting of a position, which is usually easier to make. The adjustment may, moreover, be easily checked at any time during an exposure, any systematic change noted and corrected by a slight adjustment of the pressure.

For lines which are either so weak or of such a wavelength that visual setting is impossible, one exposure is sufficient for determining the fractional part of Δ by which the line is displaced from the fringes; the air pressure can then be changed by the amount necessary to establish phase coincidence.

In these tests it is preferable for the light from the source of the line spectrum not to pass through the etalons. With external mounting this can be achieved either by means of a system of mirrors providing a by-pass for the light, or by the use of a subsidiary source. The figure shows an exposure of the heterochromatic fringes in coincidence with the line Cd II λ4416 from a commercial cadmium lamp.



Phase adjustment for Cd II λ 4416. $D = 1{\cdot}493$ cm, $m = 2$, $d = 0{\cdot}033$ cm.

The phase adjustment can also be made indirectly, by means of any suitably chosen visible line as reference line. The heterochromatic fringes offer a simple means of establishing the difference in the air pressures required to produce phase coincidence for either of two spectral lines. Once the relation with the line under investigation has been established by preliminary tests it is only necessary to adjust the phase coincidence for the reference line and then to change the pressure by the known amount. The frequent use of the source for adjustments is thus avoided; this is of advantage when the source contains rare substances or has a short life for other reasons.

In accurate measurements with double etalons it is sometimes necessary to know the suppression ratio, i.e. the ratio of the intensity of the suppressed to that of the unsuppressed fringes. This can be determined by bringing the double etalon into phase coincidence with a line which possesses no structure and has a similar wavelength, such as one of the lines emitted by a mercury isotope lamp. The mercury line λ4358 proved very suitable for use with the Cd II line λ4416.

Finally, a deliberate choice of the value of *d* can often be useful for reducing the intensity of very strong neighbouring lines which would otherwise cause stray light. Even with a chance value of *d* this action of the double etalon as interference filter is frequently very useful.

REFERENCES

JACKSON, D. A., and KUHN, H., 1938, *Proc. Roy. Soc.* A, **167,** 205.

KUHN, H., 1951, *Rep. Progr. Phys.*, **14,** 78 (London : Physical Society).

KUHN, H., and RAMSDEN, S. A., 1956, *Proc. Roy. Soc.* A, **237,** 485.

Some Elastic Properties of Organic Glasses in the Transition Region from Solid to Liquid at Ultrasonic Frequencies

By SHEILA M. CRAWFORD

Imperial Chemical Industries Limited, Akers Research Laboratories, Welwyn, Herts.

MS. received 10*th May* 1956, *and in final form* 26*th June* 1956

Abstract. Measurements have been made of the dynamic shear properties of four organic glasses at a frequency of 4 megacycles per second and over a wide temperature range covering the softening region. The behaviour of the shear modulus is in accord with the chemical nature of these materials. Correlation of these results with measurements of the velocity of longitudinal waves at the same frequency suggests that there is a relaxation in the bulk modulus between the solid and liquid state, as has been reported by Kuhn and Vielhauer, but at a somewhat lower temperature than the relaxation in the shear modulus.

§ 1. Introduction

THE dynamic shear properties of two organic glasses at different temperatures and at frequencies between 10^{-4} c/s and 10^4 c/s have already been investigated (Benbow 1954). The purpose of this work was to extend the frequency range covered in the ultrasonic region and the temperature into and beyond the transition region from the solid to the liquid state.

The glasses used were 2′-hydroxy-2 : 4 : 4 : 6 : 5′-pentamethylflavan (Baker, Curtis and McOmie 1952), 2-phenyl-3-p-tolylindan-1-one (Koelsch 1934) and sucrose octa-acetate. Sucrose octa-acetate is more brittle than the flavan compound at room temperature, with a viscosity of 10^{12} poises at 20°c compared with 10^8 poises, and it can be expected to show more intermolecular association since it contains several polar groups. The indanone compound is a red glass and has a viscosity intermediate between that of sucrose octa-acetate and the flavan compound at room temperature. Experiments were also carried out on pitch in order to observe the effect on mechanical behaviour of increasing molecular complexity.

In order to investigate the shear properties, it was decided to adopt a pulse method described by Mason, Baker, McSkimin and Heiss (1949), in which shear waves were generated in a quartz bar and reflected from one face of the bar with and without a layer of the glass adhering to it. The phase change and reduction in amplitude of the reflected wave which results from the presence of the glass were measured and from these it was possible to calculate the shear modulus, using the theory developed by O'Neil (1949). The measurements were carried out on the different glasses at a frequency of 4 megacycles per second and a wide temperature range was covered including, in one case, the transition from a solid of viscosity 10^{10} poises to a liquid of viscosity 10 poises.

Since many of the glasses used so far have been transparent, it was hoped that it might be possible to obtain their elastic constants using the optical diffraction method adopted by Schaefer and Bergmann (1948) for silicate glasses. It was found to be impossible, however, to obtain any diffraction effects due to the shear waves, and it was decided to use the optical diffraction effect merely to follow the variation with temperature of the velocity of the longitudinal waves. As far as possible, these results have been correlated with those obtained from the shear measurements to give values for the bulk modulus and Poisson's ratio.

§ 2. EXPERIMENTAL

In their original experiments, Mason and co-workers obtained the shear impedance of several liquids at ultrasonic frequencies by measuring the change in amplitude and phase of the reflected pulse when a layer of liquid was placed on the end of a quartz rod on which the shear pulses were incident normally. They found, however, that the impedances were only about 10% of that of the quartz so that the amplitude and phase were very little affected. To enhance the effect, therefore, the shear waves were reflected at a low angle of incidence and it is this modification which has been used in these experiments.

Two similar bars of transparent fused quartz (figure 1) were used and identical gold-sputtered Y-cut crystals (T) were attached to the ends of the bars with their diad axes parallel to the reflecting surface so that the shear would be in the plane of this surface. A melt adhesive M. 136 supplied by the B.B. Chemical Company was used to fix the crystals. The adhesive had the advantage that it remained solid up to 100°c and so transmitted the shear pulses in the temperature range required, but since it melts at 180°c, the crystals could easily be adjusted or removed. The pulses were provided by a double-tetrode oscillator arranged to be self-pulsing and to give pulses of 10 microseconds length.

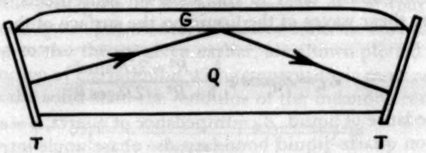


Figure 1. Apparatus for shear measurements. Q, quartz bar; G, organic glass; T, transmitting and receiving crystals.

For the reflection coefficient measurements one bar only was used, and the output from the receiving crystal was displayed on an oscilloscope and photographed with and without the organic glass adhering to the reflecting surface. The ratio between the amplitudes of the first reflected pulse with and without the glass on the surface gave the reflection coefficient.

The change of phase produced in the reflected pulse when the reflecting surface was covered with glass was measured by a null method. Both bars were used and the connections to the crystals were arranged so that the received pulses with no glass on either bar balanced each other out. The reflecting surface of one of the bars was then covered with the glass and an attenuator which did not alter the phase appreciably was used to equalize the amplitudes again. The pulses were brought into opposition again by inserting into the receiving circuits different lengths of coaxial cable which would alter the phase by a known amount. In this way the phase change introduced by the glass could be directly determined and at best it was possible to make measurements to ± 1°.

The variation with temperature of the density of these glasses had been previously determined so that, as shown in the next section, the dynamic shear elasticities could be calculated from the measurements of the reflection coefficient and phase change at different temperatures.

The same oscillator was used in the diffraction experiments except that it was adjusted to give a continuous wave instead of pulses. The vibrations from the quartz crystals vibrating in their fundamental thickness mode were transmitted to the organic glass in a silicate cell through a thin layer of oil. The optical arrangement was the same as that adopted by Debye and Sears (1932) and Lucas and Biquard (1932) using a mercury vapour lamp fitted with a green filter to give a monochromatic beam of wavelength $\lambda = 5461$ Å.

The diffraction images produced by the ultrasonic 'grating' were photographed and the distance d_p between the pth diffraction image and the direct image was measured. The frequency N of the longitudinal waves was determined with a wavemeter so that, knowing the focal length F of the lens focusing the pattern, the velocity V of the waves could be calculated from the simple diffraction formula

$$\sin\theta \simeq \frac{d_p}{F} = \frac{pN\lambda}{V} \quad \ldots\ldots(1)$$

§ 3. THEORY

It was shown by Mason and co-workers (1949) from the work of O'Neil (1949) that, for the arrangement used in the shear measurements, the characteristic impedance to the shear waves of the liquid on the surface of the quartz was given by

$$Z_L = Z_Q \cos\phi \, \frac{1 - R^2 + 2iR\sin\theta}{1 + R^2 + 2R\cos\theta} \quad \ldots\ldots(2)$$

where Z_L = impedance of liquid, Z_Q = impedance of quartz, ϕ = angle of incidence of shear waves on quartz–liquid boundary, θ = phase angle introduced by liquid and R = relative reduction in amplitude of the waves due to the liquid.

In general, the impedance is a complex quantity and for shear waves is given by

$$Z = [\rho(G' + iG'')]^{1/2} \quad \ldots\ldots(3)$$

where ρ = density of medium, G' = real part of the complex shear modulus and G'' = imaginary part of the complex modulus.

The losses are negligibly small in quartz so that Z_Q is a real quantity and hence the real and imaginary parts of the shear modulus of the organic glass can be obtained from (2) and (3).

$$G' = \frac{Z_Q{}^2 \cos^2\phi}{\rho} \left\{ \frac{(1 - R^2)^2 - 4R^2 \sin^2\theta}{(1 + R^2 + 2R\cos\theta)^2} \right\} \quad \ldots\ldots(4)$$

$$G'' = \frac{Z_Q{}^2 \cos^2\phi}{\rho} \left\{ \frac{4R(1 - R^2)\sin\theta}{(1 + R^2 + 2R\cos\theta)^2} \right\}. \quad \ldots\ldots(5)$$

Z_Q is known and ϕ in this experiment was 79°. ρ, R and θ were determined experimentally so that both G' and G'' could be evaluated.

The real part of the velocity of the longitudinal waves in the material is given by

$$V_S = [(K' + 4G'/3)/\rho]^{1/2} \qquad \ldots\ldots(6)$$

where K' and G' are respectively the real parts of the dynamic bulk and shear moduli and ρ is the density of the medium. Since ρ and G' have been evaluated, measurements of the velocity of the longitudinal waves make it possible to determine K'.

The bulk and shear moduli are related to each other through Poisson's ratio whose real part ν is given by

$$\nu = \frac{3K' - 2G'}{6K' + 2G'}. \qquad \ldots\ldots(7)$$

§ 4. RESULTS

The results for the shear experiments with the flavan compound are shown in figure 2, where the relative reduction in amplitude and the phase shift of the reflected waves produced by the presence of the glass on the quartz are plotted against temperature. The rapid rise in the amplitude ratio and the peak in the phase shift curve are associated with the softening of the material. The curves for the other materials are similar except that the peak is somewhat broader for the tolylindanone compound and still more so for the sucrose octa-acetate and pitch. There is a corresponding spread of the temperature range over which the amplitude ratio has its most rapid rate of rise and this rate is considerably less in the case of sucrose octa-acetate and pitch. The peaks in the phase shift curves all occur at temperatures where the viscosity, as measured by a simple Couette type of viscometer, is of the order of 1000 poises.

In figure 3 the real parts of the shear moduli of the four substances at a frequency of 4 megacycles per second, calculated from the smoothed experimental results according to the theory given earlier, are shown plotted against temperature. The relaxation of the modulus with temperature is most rapid in the flavan compound. In the solid state the modulus of the indanone compound is rather

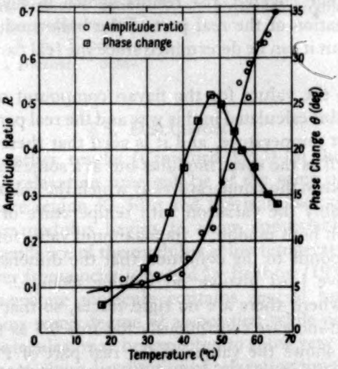


Figure 2. Amplitude ratio and phase change versus temperature for 2′-hydroxy-2 : 4 : 4 : 6 : 5′-pentamethylflavan at 4 Mc/s.

more temperature dependent than that of the flavan compound but less so than that of sucrose octa-acetate. It is very temperature dependent in pitch and shows an almost continuous relaxation into the liquid state. It is of particular interest to observe that the shear elasticity of the flavan compound is still of the order of 10^9 dyn cm^{-2} when it is a liquid of viscosity 10 poises.

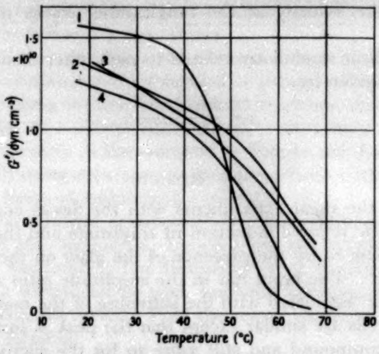


Figure 3. Real part of shear modulus versus temperature at 4 Mc/s for 1, 2′-hydroxy-2 : 4 : 4 : 6 : 5′-pentamethylflavan, 2, sucrose octa-acetate, 3, pitch, 4, 2-phenyl-3-p-tolylindan-1-one.

Figure 4 shows the results obtained from the diffraction experiments for the velocity of longitudinal waves of frequency 4 megacycles per second in sucrose octa-acetate and the flavan compound. It was not possible to make any measurements for the latter in the region shown by the broken line because the diffraction images became too faint, and it was not possible to work at sufficiently high temperatures to observe diffraction effects in liquid sucrose octa-acetate. Since the flavan compound has been shown to be capable of supporting shear throughout the temperature range covered, the results shown in figure 4 do not allow of the direct determination of the real part of the bulk modulus from the velocity in the liquid state, but it can be determined since the real part of the shear modulus is known.

Figure 5 shows the values for the flavan compound of the real part of the dynamic bulk modulus calculated in this way and the real part of the shear modulus both plotted against temperature, and it is seen that there is a relaxation in the bulk modulus as well as the shear modulus but at a somewhat lower temperature. This difference between the temperatures at which the relaxation processes occur will affect considerably the variation with temperature of Poisson's ratio since the latter depends on both moduli. Its maximum value for an isotropic material is 0·5 which corresponds to the condition that the dimensional changes accompanying an extensive load involve no volume change. This condition only applies in a liquid where there are no rigid forces, so that it might be expected that the value would increase continuously towards 0·5 as the material softened. Figure 6, however, shows the values of the real part of Poisson's ratio plotted against temperature for both sucrose octa-acetate and the flavan compound and it is seen that in the case of the flavan compound there is a marked decrease with increasing temperature at first, which gives rise to a minimum before the ratio increases and tends towards the maximum value of 0·5. The effect is less marked in the sucrose octa-acetate.

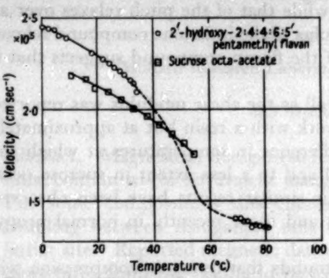


Figure 4. Velocity of longitudinal waves versus temperature at 4 Mc/s in 2′-hydroxy-2 : 4 : 4 : 6 : 5′-pentamethylflavan and sucrose octa-acetate.

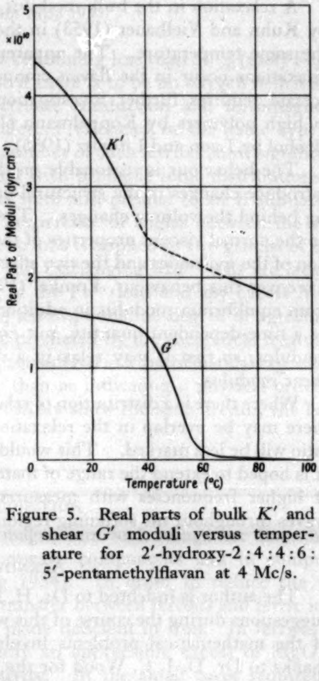


Figure 5. Real parts of bulk K' and shear G' moduli versus temperature for 2′-hydroxy-2 : 4 : 4 : 6 : 5′-pentamethylflavan at 4 Mc/s.

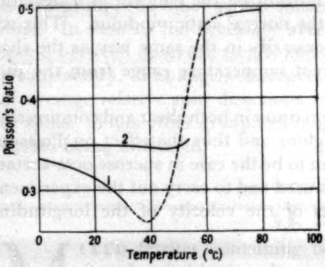


Figure 6. Real part of Poisson's ratio at 4 Mc/s versus temperature for 1, 2′-hydroxy-2 : 4 : 4 : 6 : 5′-pentamethylflavan, 2, sucrose octa-acetate.

§ 5. Discussion

The trend of the earlier results (Crawford 1953) on the flavan compound would suggest that the relaxation time will be of the same order as the applied stress in the temperature region 40–50°c and this is, in fact, the region where the real part of the shear modulus begins to show appreciable relaxation. The small temperature dependence of the modulus at low temperatures is in agreement with the results at lower frequencies reported by Benbow (1954).

The sucrose octa-acetate molecule contains several polar groups and is therefore likely to show some degree of association while pitch, which contains many such types of molecules, may be expected to show very marked association. The effect of this will be to give a distribution of relaxation times at a given temperature rather than a single value as appears to be the case in the flavan compound which is not expected to show much association. It might be anticipated, therefore, that the temperature region over which the maximum relaxation in the modulus occurs will be broadened according to this distribution and so it is not surprising that the modulus of sucrose octa-acetate does relax over a wider temperature range than that of the flavan compound, while that of the pitch relaxes over an even wider range. The fact that the modulus of the indanone compound relaxes over a somewhat wider range than that of the flavan compound suggests that it might be slightly associated.

A relaxation in the bulk modulus as well as the shear modulus was reported by Kuhn and Vielhauer (1953) in their work with a resin but at approximately the same temperature. The apparent difference in temperatures at which the relaxations occur in the flavan compound and to a less extent in sucrose octa-acetate requires further investigation, but similar results have been observed in high polymers by Koppelmann (1955) and more recently in normal propyl alcohol by Lyon and Litovitz (1956).

The behaviour is reasonable on the grounds that dynamic compression will introduce changes in the structural order of the medium which, in general, will lag behind the volume changes. These structural alterations will depend partly on the normal viscous properties of the material and partly on changes in orientation of the molecules and the two effects may not have the same time dependence. To cover this behaviour, Frenkel (1946) introduced the concept of a deviation from equilibrium modulus in addition to the normal static modulus. This will be a time-dependent quantity, but not necessarily in the same way as the shear modulus, so that it may relax in a different temperature range from the pure shear modulus.

Where there is a distribution of relaxation times in both shear and compression, there may be overlap in the relaxation regions and thus the effect on Poisson's ratio will be less marked. This would seem to be the case in sucrose octa-acetate. It is hoped to extend the range of materials used and to carry out the experiments at higher frequencies with measurements of the velocity of the longitudinal waves throughout the softening region.

ACKNOWLEDGMENTS

The author is indebted to Dr. H. Kolsky for his encouragement and valuable suggestions during the course of this work and to Dr. F. C. Roesler for discussion of the mathematical problems involved. She would also like to express her thanks to Dr. D. J. C. Wood for the preparation of the glasses and to Mr. J. J. Benbow for access to his results.

REFERENCES

BAKER, W., CURTIS, R. F., and MCOMIE, J. F. W., 1952, *J. Chem. Soc.*, 1774.
BENBOW, J. J., 1954, *Proc. Phys. Soc.* B, **67**, 120.
CRAWFORD, S. M., 1953, *Proc. Phys. Soc.* B, **66**, 953.
DEBYE, P., and SEARS, F. W., 1932, *Proc. Nat. Acad. Sci., Wash.*, **18**, 409.
FRENKEL, J., 1946, *Kinetic Theory of Liquids* (Oxford : Clarendon Press).
KOELSCH, C. F., 1934, *J. Amer. Chem. Soc.*, **56**, 1337.
KOPPELMANN, J., 1955, *Kolloidzschr.*, **144**, 12.
KUHN, W., and VIELHAUER, S., 1953, *Z. phys. Chem.*, **202**, 161.
LUCAS, R., and BIQUARD, P., 1932, *C. R. Acad. Sci., Paris.*, **194**, 2132.
LYON, T., and LITOVITZ, T. A., 1956, *J. Appl. Phys.*, **27**, 179.
MASON, W. P., BAKER, W. O., MCSKIMIN, H. J., and HEISS, J. H., 1949, *Phys. Rev.*, **75**, 936.
O'NEIL, H. T., 1949, *Phys. Rev.*, **75**, 928.
SCHAEFER, C., and BERGMANN, L., 1948, *Ann. Phys., Lpz.*, **3**, 72.

Angles between Magnetic Spin Directions in Iron-Deficient Magnesium Manganese Ferrites

By W. P. OSMOND

Mullard Research Laboratories, Salfords, nr. Redhill, Surrey

MS. received 24th May 1956

Abstract. Magnesium manganese ferrites containing less than 50% (mol) Fe_2O_3 must contain tri- or tetravalent manganese if there is to be no oxygen deficiency. For some compositions this entails electronic conduction and ferromagnetic coupling between manganese ions of different valencies in the same type of lattice site. Reported magnetic data on a number of such ferrites show significant departures from the results to be expected with probable ionic distributions in the spinel lattice if all A sub-lattice spins were anti-parallel to all B sub-lattice spins. Existing theoretical work on the occurrence of angles between the spin directions of two different kinds of magnetic ions is applied to two groups of ions in one or both sub-lattices. The Mn^{3+} and exchanging Mn^{2+} ions are treated as one group, the other group consisting of the Fe^{3+} ions and any excess Mn^{2+} ions. In view of the necessary averaging assumptions the degree of agreement found between observed trends and those predicted by the theoretical treatment is satisfactory. The angles concept is here considered as a mathematical expression of average relative spin directions rather than as indicating a physical ordering of spins. Thus other groupings might possibly show the same trends, the basic causes being unaltered.

§ 1. Introduction

Mixed ferrites containing both magnesium and manganese have been found suitable for use in microwave components and in computer circuits (cf. Albers-Schoenberg 1954). In order to reduce the risk of lowered resistivity caused by electron transfer between ferrous and ferric ions, the compositions are always in practice made deficient in iron. In ferrites for microwave applications there is frequently an appreciable excess of MgO, in order still further to increase the resistivity. In the small cores required to exhibit rectangular hysteresis loops in magnetic storage systems the magnesium manganese ratio is usually relatively less, since this is found to result in less porosity and better hysteresis characteristics.

The general chemical formula for such ferrites is $Mg_pMn_qFe_rO_4$ with $p+q+r=3$, and, for the iron-deficient compositions considered here, $r<2$. When the manufacturing precautions have been such as to ensure that only ferric iron is present, some of the manganese must be trivalent or tetravalent if there is to be minimum oxygen deficiency. In 'microwave' ferrites containing more magnesium than is represented symbolically by $(MgFe_2O_4)_{1-q}(Mg_2MnO_4)_q$, $p>1+q$ and there must always be some oxygen deficiency even though all the manganese is Mn^{IV}. However, for 'square-loop' ferrites of smaller relative magnesium content, firing in an atmosphere containing oxygen can ensure complete oxygen compensation, the manganese valency then adjusting itself to the requisite value. When Mn ions of different valencies appear in the same type of crystal lattice site, lowered resistivity is to be expected. In this paper the electrical and magnetic properties reported by Economos (1954) for thirteen such square-loop ferrites are examined for evidence of the probable distribution of the various ions in the crystal lattice and of the relative strengths of their mutual interactions.

§ 2. RANGE OF COMPOSITIONS STUDIED

The ferrites of which the reported properties are to be discussed were all fired in flowing air at a temperature below that at which a reduction in saturation and remanent magnetization (attributable to the formation of ferrous iron) was found to set in. Hence the conditions mentioned in § 1 seem to have been very fairly well observed. The compositions, expressed as mol. percentages of constituent oxides, lie in the ranges Fe_2O_3 25·3–40·3, MgO 25·6–40·6, Mn^xO 26·6–44·1, where x has been inserted here to indicate the average manganese valency for no oxygen deficiency. From these percentages the corresponding values of the suffixes p, q in the chemical formula can readily be calculated, and also the oxygen deficiency $s=(p+q-1)/2$ which would result with Mn^{II} only. All these ferrites were found to be single-phased cubic spinels, with lattice parameter a varying from 8·43 to 8·47 Å.

Seven of the compositions contain less MgO than Fe_2O_3. They can, therefore, be represented as mixtures of $MgFe_2O_4$, $MnFe_2O_4$ and Mn_3O_4, all of which compounds are known to exist. The remaining six ferrites, with more magnesium and manganese than simple $MgFe_2O_4$–Mn_3O_4 mixtures, but less magnesium than is represented by $p=1$, can be treated as mixtures of these two compounds and $MgMn_2O_4$. The existence of the latter has not been established with certainty, but the formula provides a convenient symbol to show that with full oxygen content the extra manganese must be trivalent.

§ 3. THE SYSTEM $MgFe_2O_4$—$MnFe_2O_4$—Mn_3O_4

The composition formula for the seven ferrites in which the mol. proportion of MgO is less than that of Fe_2O_3 (i.e. $p<r/2$) can be written as $(MgFe_2O_4)_p(MnFe_2O_4)_w(Mn_3O_4)_s$ with $p+w+s=1$. Here $w=\frac{1}{2}r-p$ and $s=(q-\frac{1}{2}r+p)/3=(p+q-1)/2$, the amount of oxygen compensation. The first two constituents are simple spinel ferrites for which the ionic distributions are well known (cf. Gorter 1954), while Mn_3O_4 is a tetragonal spinel for which Goodenough and Loeb (1955) argue convincingly that there are two Mn^{3+} ions in the octahedral, B, sites and one Mn^{2+} ion in the tetrahedral, A, site. In these mixtures we might then expect to find the A sites occupied in the proportions $0{\cdot}11pMg^{2+}$, $(0{\cdot}89p+w)Fe^{3+}$ and $s\,Mn^{2+}$, and the B sites in the proportions $0{\cdot}89pMg^{2+}$, $(1{\cdot}11p+w)Fe^{3+}$, wMn^{2+} and $2sMn^{3+}$. If so, there should be increasing electron transfer between the Mn^{2+} and Mn^{3+} ions in B sites as w increases, so long as the latter remains less than $2s$. Support is given to this proposed distribution by the reported resistivity values. Those for the three compositions with $w\sim 0{\cdot}2$ are considerably lower than those for the four with $w\simeq 0{\cdot}05$, the values for both sets increasing as expected with larger content of $MgFe_2O_4$, for which $\rho\sim 10^9$ ohm cm compared with 10^7 for Mn_3O_4.

With this distribution we can estimate the probable net magnetic moment in Bohr magnetons $\boldsymbol{\beta}$ per formula unit at 0°K, assuming the simple model given by Néel (1948), with the spins of the magnetic ions in B sites all parallel to each other and anti-parallel to the spins of all those in A sites. For this purpose we know (Gorter 1954) that the Landé splitting factor g is 2 for the Fe^{3+} and Mn^{2+} ions, while Jonker and van Santen (1950) have given 1·99 for the Mn^{3+} ion. We thus find $5\boldsymbol{\beta}$ for the saturation moment of both Fe^{3+} and Mn^{2+} and $4\boldsymbol{\beta}$ for that of Mn^{3+}. The suggested distribution then gives for the net moment per formula unit $Mg_pMn_qFe_rO_4$ at 0°K, $n_\beta = 3 + 2w - 1{\cdot}9p$. For constant w this increases with decreasing p, and the same trend would be shown if the Mg ions were not distributed exactly as in $MgFe_2O_4$.

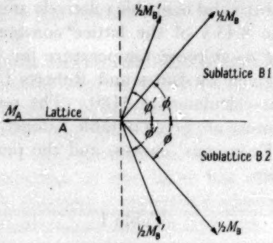
Study of the reported values of maximum flux density B_m at a constant magnetizing field of 25 oersteds (which in view of the coercivities of 1 or 2 oersteds may reasonably be considered to saturate the material) shows that this general trend of variation with composition is followed fairly well at ordinary temperatures. However, for close comparison it is essential to relate theoretical and experimental values to the same temperature. In default of any other data we must, therefore, calculate an approximate value of the ratio of saturation intrinsic induction $4\pi M_s$ in gauss at room temperature to n_β in units of $\boldsymbol{\beta}$ at 0°K, which should be reasonably valid over this relatively small range of compositions. Using the average value 8·45 Å of the lattice constant, and the mean of the percentage reductions of n_β at room temperature for $MgFe_2O_4$ and $MnFe_2O_4$, estimated from curves given by Jones and Roberts (1952) and Gorter (1954) respectively, this ratio is calculated as 1180. The resulting theoretical values of $4\pi M_s$ at room temperature are given in table 1, together with the experimental figures estimated from Economos' curves, and the percentage differences Δ of the latter from the former.

Table 1

Ferrite	p ($MgFe_2O_4$)	q	w ($MnFe_2O_4$)	s (Mn_3O_4)	$4\pi M_s$ (gauss) theor.	$4\pi M_s$ (gauss) exptl.	Δ (%)
G	0·77	0·58	0·055	0·175	1945	1675	−13·9
A	0·73	0·70	0·055	0·215	2035	1685	−17·2
B	0·69	0·83	0·05	0·26	2110	1735	−17·8
K	0·65	0·95	0·05	0·30	2200	1680	−23·6
H	0·66	0·62	0·20	0·14	2530	1865	−26·3
I	0·61	0·74	0·215	0·175	2680	1875	−30·0
J	0·57	0·86	0·215	0·215	2770	1935	−30·1

In spite of the errors inherent in the methods of estimation and calculation, which include no allowance for possible porosity, there appears to be a definite trend in each set of compositions for a larger deficiency of magnetic moment relative to the theoretical value with increasing number ($2s$) of Mn^{3+} ions. Moreover, the deficiencies are largest in the second set, with relatively fewer Mn^{3+} but considerably more Mn^{2+} ions w together in B sites. Since the theoretical net moment is parallel to that of the B sub-lattice, all B spins being then assumed anti-parallel to all A spins, a natural inference is that in fact fewer B spins are absolutely parallel when Mn^{3+} ions are found in these sites, and fewer still when they are accompanied by Mn^{2+} ions. The former idea accords with the contention of Goodenough and Loeb (1955) that the A–B anti-parallel interaction is much reduced when the B ion is Mn^{3+}, since the latter tends to form square covalent bonds with its oxygen neighbours. The second inference agrees with the theory of double exchange (Zener 1951), according to which the electron transfer between similarly situated ions of the same substance but different valencies leads to parallelism of their spins.

The combination of these two effects thus leads to the idea of an angle between the spin directions of two groups of ions in the B sub-lattice, one consisting of the Fe^{3+} ions subjected to a strong A–B interaction, the other comprising the Mn^{3+} and Mn^{2+} ions undergoing double exchange and subjected to a weak A–B interaction. Distinguishing the total moments of these groups as M_B and M_B' respectively, and denoting by M_A the grouped moment of the magnetic ions in A sites (Fe^{3+} and Mn^{2+}, which on account of their similarity will be expected to have mutually parallel spins as in the B sub-lattice of $MnFe_2O_4$), the simplest geometrical representation of this idea is the symmetrical array shown in the figure. This subdivision into two sub-lattices B1 and B2 is the arrangement suggested by Niessen (1953) for a ferrite having two kinds of magnetic ions in B sites, one kind being the same as the single type in A sites.



Symmetrical arrangement of partial magnetic moments for two different groups of ions in the B spinel sub-lattice.

The development of the molecular field treatment as indicated by Niessen is a lengthy process involving several averaging assumptions. This has been carried out (Osmond 1956), using the fact that in the present instance all angles would be expected to vanish with M_B' (no Mn^{3+} ions). It is then found that the average relative strength of the anti-parallel B–B interactions for a small range of compositions varies directly as the number of magnetic A ions and inversely as the number of B ions of the same type. In both sets of compositions listed in table 1 the suggested number $0{\cdot}89p+w+s$ of magnetic A ions increases continuously while the number $1{\cdot}11p+w$ of similar B ions decreases continuously. The latter would still be true if all the Mg ions were in B sites, while the number of A ions (all magnetic) would remain constant. The modified Niessen treatment then predicts a continuous increase in the relative strength of the B–B′ interactions, hence in the size of the angles ϕ and ϕ' shown in the figure and in the relative reduction of the net moment. Thus this method, regarded simply as a convenient mathematical tool for the estimation of the average effects of definite physical differences, does in this instance give qualitative agreement with the apparent trend of the reported magnetic measurements. At the same time the suggested ionic distribution fits the resistivity trends.

§ 4. THE SYSTEM $MgMn_2O_4$—$MgFe_2O_4$—Mn_3O_4

In the six ferrites having a larger proportion of MgO than of Fe_2O_3 we have $1 > p > r/2$. The composition formula can then be expressed as

$$(MgMn_2O_4)_k(MgFe_2O_4)_l(Mn_3O_4)_m$$

with $k+l+m=1$. Here $k=p-r/2$, $l=r/2$, and $m=(q+r-2p)/3=1-p$. The $2k\,Mn^{3+}$ ions associated with the first constituent might be expected to seek B sites as in Mn_3O_4. If so, the total of $2(m+k)\,Mn^{3+}$ ions in these sites might well lead to a tetragonal phase when the proportion of $MgFe_2O_4$ is small; also the resistivity should be high, there being only one kind of manganese ion in each type of site. However, all the compositions in this group are reported to be single-phased cubic spinels. Moreover, the reported resistivities, increasing as expected with the content l of $MgFe_2O_4$ and constant k, are lower for the pair of compositions with $k \simeq 0{\cdot}3$ then for the four with $k \simeq 0{\cdot}12$. The inference is that some of these $2k\,Mn^{3+}$ ions are to be found in A sites as well as the $m\,Mn^{2+}$ ions expected there. This suggests that the $MgMn_2O_4$ should here be treated like an inverse spinel $Mn^{3+}[MgMn^{3+}]O_4$ as in the known example of $MgIn_2O_4$ (Barth and Posnjak 1932).

On this basis we should expect to find the A sites to be occupied in the proportions $0{\cdot}11l\,Mg^{2+}$, $0{\cdot}89l\,Fe^{3+}$, $m\,Mn^{2+}$ and $k\,Mn^{3+}$, and the B sites in the proportions $(0{\cdot}89l+k)Mg^{2+}$, $1{\cdot}11l\,Fe^{3+}$ and $(2m+k)Mn^{3+}$. If the simple Néel law were then followed, the net moment per formula unit at 0°K would be $n_\beta = 1{\cdot}1l+3m = 1{\cdot}1(1-k)+1{\cdot}9m$. This value increases with increasing m, both for constant l and for constant k. The former trend is apparent from the reported values of $4\pi M_s$ but not the latter. However, no other distribution of the $MgMn_2O_4$ constituent has been found to give better agreement, and none fits the resistivity trends so well. The reported values of $4\pi M_s$, those calculated from the theoretical formula, and the percentage differences of the former from the latter are given in table 2.

Table 2

Ferrite	p	q	l	k	m	$4\pi M_s$ (gauss)		$\Delta(\%)$
			($MgFe_2O_4$)	($MgMn_2O_4$)	(Mn_3O_4)	theor.	exptl.	
F	0·86	0·66	0·74	0·12	0·14	1455	1735	+19·2
O	0·82	0·78	0·70	0·12	0·18	1545	1675	+8·4
C	0·78	0·92	0·65	0·13	0·22	1620	1645	+1·5
L	0·73	1·06	0·605	0·125	0·27	1740	1645	−5·5
E	0·95	0·75	0·65	0·30	0·05	1020	1205	+18·1
D	0·91	0·88	0·605	0·305	0·09	1105	1185	+7·2

The most striking feature of this table is that for all but one of the six ferrites the reported values of $4\pi M_s$ are *higher* than those given by theory. However, for each set the relative excess decreases with increasing number $(2m+k)$ of Mn^{3+} ions expected to be in B sites, becoming a deficit for the composition L in which this number is a maximum. The apparent trend is thus in the same sense as that shown by the first group of ferrites. Hence it could be described as the result of an increasing angle between the direction of the Mn^{3+} spins and that of the Fe^{3+} spins in B sites. As before, the physical idea underlying this mathematical concept is that the anti-parallel A–B interaction is weaker with Mn^{3+} than with Fe^{3+} ions in B sites. Again, measured values above the theoretical figures could be attributed to relatively large reductions in the total A sub-lattice moment. This can similarly be expressed as the result of angles between spins in this sub-lattice. In support of this idea we here expect to find $k\,Mn^{3+}$ ions in A sites, unlikely to form strong covalent bonds, but likely to be coupled by double exchange with parallel spins to some or all of the $m\,Mn^{2+}$ ions also expected there. Thus we can now picture groups of ions, of the same two types, in both sub-lattices.

The symmetrical array of partial moments shown in the figure for two B sub-lattices can be extended to two A sub-lattices as well, with different angles, to give a geometrical representation of reduced net moments in both types of site. The corresponding molecular field treatment has been carried out (Osmond 1956) with the same averaging assumptions as in the previous case. It is then found that the average relative strength of the anti-parallel A–A′ interactions for a small range of compositions varies inversely as the number of A ions and directly as the number of B ions in the groups subjected to strongest A–B interactions (Fe^{3+} ion groups here). The average relative strength of the B–B′ interactions varies in the opposite way, as in the previous case. In both sets of compositions in table 2 the suggested number $1{\cdot}11l$ of Fe^{3+} ions in B sites decreases continuously while the number (different in the two sets) of grouped Fe^{3+} and Mn^{2+} ions in A sites changes little. Here also the same trend would be observed if all the Mg ions were in B sites. The modified Niessen treatment then predicts a gradual increase in the relative strength of the B–B′ interactions and decrease in that of the A–A′ interactions, hence in the size of the angles and relative reductions of the net moments in the respective sub-lattices. Again this prediction qualitatively agrees with the apparent trend. In particular, when the number of ions of the Fe^{3+} group in A sites is greater than the number of similar B ions this treatment predicts that the reduction in the net B moment may well be greater than that in the net A moment. With the suggested distribution ferrite L is the only one with this preponderance of the A group, and it is also the only one with this apparent relative reduction of the anti-parallel net moments. However, this complete agreement is possibly rather fortuitous in view of the small number of compositions studied and the uncertainties of the actual ionic distribution. For instance, if all the Mg ions were in B sites the A group would be larger than the B group for all four ferrites of the first set, the relative preponderance increasing to a maximum for composition L. For both distributions, therefore, the apparent trends are fitted by this angles treatment.

§ 5. Conclusion

In this study of a limited range of iron-deficient magnesium manganese ferrites a degree of qualitative agreement has been found between the apparent trends of the variation with composition of room temperature measurements of saturation magnetization and those to be expected on two main theoretical assumptions. These are (i) that the anti-parallel interactions between Fe^{3+} and

Mn^{3+} ions in different types of spinel lattice site are relatively weak, (ii) that ferromagnetic coupling by double exchange takes place between similarly situated Mn^{2+} and Mn^{3+} ions. The assumed ionic distributions in the lattice are consistent with electrical resistivity trends and known distributions in simpler compositions. In view of the uncertainties of the actual distributions and of the temperature variation of magnetic moments the fact that there is any agreement at all lends weight to these theoretical ideas.

The reduction of the net moment of any sub-lattice containing Mn^{3+} ions, predicted by these concepts, is conveniently described as caused by an angle between the spin directions of two dissimilar groups of ions one of which contains the Mn^{3+} ions. This idea of angles is used here as a mathematical tool for expressing reduced net moments at room temperature, and not essentially as indicating a physical ordering of spins in equivalent crystallographic sub-lattices. The ionic grouping adopted, although the most obvious, is therefore not necessarily the only arrangement that would show the same trends by an angles treatment. In fact, it is shown in a paper to be presented by the author at the I.E.E. Convention on Ferrites that the theory of semi-covalent exchange (Goodenough and Loeb 1955), which can account fully for the known magnetic properties of simple ferrites, can explain qualitatively the trends observed in these mixed ferrites without any specific anti-parallel A–A and B–B interactions, which are the basis of the standard concept of angles. With this new idea of exchange, the effect of an isolated pair of double exchanging manganese ions in B sites is to make the spins of four Fe^{3+} ions less likely to be parallel to the net moment of this sub-lattice that those of all the other B ions, including the manganese ions, and to cause some departure from the complete parallelism of all A spins also.

ACKNOWLEDGMENTS

I wish to thank Mr. P. E. Trier, Manager of the Mullard Research Laboratories, for permission to publish this paper, and Dr. K. Hoselitz, Head of the Solid State Physics Group, for his interest and encouragement and for helpful criticisms of the original manuscript.

REFERENCES

ALBERS-SCHOENBERG, E., 1954, *J. Appl. Phys.*, **25**, 152.
BARTH, T. F. W., and POSNJAK, E., 1932, *Z. Kristallogr.*, **82**, 325.
ECONOMOS, G., 1954, Laboratory for Insulation Research, M.I.T., Technical Report 78.
GOODENOUGH, J. B., and LOEB, A. L., 1955, *Phys. Rev.*, **98**, 391.
GORTER, E. W., 1954, *Philips Res. Rep.*, **9**, 295, 403.
JONES, G. O., and ROBERTS, F. F., 1952, *Proc. Phys. Soc.* B, **65**, 390.
JONKER, G. H., and VAN SANTEN, J. H., 1950, *Physica*, **16**, 337.
NÉEL, L., 1948, *Ann. Phys., Paris*, **3**, 137.
NIESSEN, K. F., 1953, *Physica*, **19**, 1085.
OSMOND, W. P., 1956, *Phil. Mag.*, **1**, 8th (Ser.), 1147.
ZENER, C., 1951, *Phys. Rev.*, **82**, 403.

The Vibrational Frequency Distribution of Graphite : I. Out-of-plane Modes of a Single Layer

By P. B. SHERRY and C. A. COULSON
Mathematical Institute, Oxford

MS. received 30*th August* 1956, *and in final form* 4*th October* 1596

Abstract. Using a potential energy function derived from molecular spectra to describe the out-of-plane vibrations of a graphite crystal, the distribution function for these vibrations is calculated for an infinite, single layer. This distribution function is different from certain other calculated distributions which use other types of potential energy function. The reason for this difference is analysed, and application is made to the contribution of these modes of vibration to the specific heat of graphite.

§ 1. Introduction

Recent experimental data on the specific heat of graphite at low temperatures (DeSorbo and Tyler 1953, 1956) has stimulated theoretical research on the determination of the lattice vibrations of this type of solid (Rosenstock 1953, Newell 1955, 1956, Yoshimori and Kitano 1956, Komatsu and Nagamiya 1951, Komatsu 1955). Determinations of the structure of graphite show that it is composed of layers of tightly packed carbon atoms. The atoms within a layer are arranged in a hexagonal net and are separated from one another by a distance of 1·42 Å. The parallel layers are separated by a distance of 3·35 Å and are arranged so that the carbon atoms of every other layer are directly over one another, while the in-between layers are rotated around one of the atoms through an angle of sixty degrees relative to the layers above and below it. This structure gives rise to a substance of a high degree of anisotropy, and it is for this reason that the variation in the specific heat of graphite with temperature would be expected to differ somewhat from the dependence predicted by the simple Debye formula.

Two methods are available to allow for the anisotropic nature of the graphite lattice. The first method (Komatsu and Nagamiya 1951, Komatsu 1955) involves a modification of the elastic model of Debye to take into account the layer-like structure of graphite. This method considers a graphite crystal as a system of thin, equally spaced elastic plates and allows for the coupling of the vibrations in each layer. Since this method pays no attention to the discrete nature of the graphite lattice, it has the same disadvantages as the conventional Debye treatment. Another more serious drawback to the method is that the distribution of vibrations must be obtained in terms of certain elastic constants which are difficult to determine from experimental data.

The second method of approach to the vibrations of the graphite lattice uses the Born–von Kármán method. In this method the discrete nature of the lattice is taken into account by solving the equations of motion for the individual carbon atoms. In order to apply the method a potential energy function must be chosen which allows the equations of motion to be solved and which is free of arbitrary constants that are not available from experimental data. In the past, workers have tried to develop the potential function in terms of simple Hooke's law forces between first, second and third nearest neighbours; but this method has the drawback that the force constants are not easily determined, nor always easily visualized.

The present paper uses the Born–von Kármán method and deals with the difficulties encountered by previous authors by the introduction of a potential function in which the various terms can be related to well-known properties of molecular bonds. The constants which appear in this potential function can be determined, to a good approximation, by the infra-red absorption spectra of compounds which are chemically similar to graphite.

With a single layer it is easy to see that the in-plane and out-of-plane vibrations may be considered quite separately, and in this paper we are concerned solely with the out-of-plane vibrations of a single layer. The in-plane vibrations of a graphite lattice have already been discussed by Yoshimori and Kitano (1956). They use a potential function which involves the change in the length of the bond joining nearest neighbours and, in addition, a term involving the change in the angle between any two bonds joining an atom with any two of its nearest neighbours.

§ 2. POTENTIAL ENERGY FUNCTION

The potential energy function used in this work was developed by Coulson and Senent (1955 a) for the study of the out-of-plane vibrations of ethylene and aromatic molecules. Later it was used with success to determine the molecular deformation of overcrowded molecules (Coulson and Senent 1955 b). The Coulson–Senent potential function divides the potential energy of each ethylenic-type region into a term involving the torsion of the bond joining the two central carbon atoms and a flapping term involving the out-of-plane motion of the two pairs of attached atoms. (In the case of ethylene these atoms would be hydrogen, but in our case of an infinite layer of carbon atoms they would be carbon.) To a first approximation the cross product terms are neglected. If this potential function is applied to a hexagonal net of bonds such as that of a layer of graphite, the potential energy for any one atom will involve three torsion terms for the bonds joining the atom with its first nearest neighbours and in addition three flapping terms involving the three pairs of nearest neighbours. In figure 1, for example, the energy associated with atom 5 and bond 5·6 is given by the formula

$$V=\tfrac{1}{2}K_1(z_1+z_2+z_6-3z_5)^2+\tfrac{1}{2}K_2(z_2-z_1-z_4+z_3)^2.$$

The first two terms are the lack-of-planarity terms while the last is the torsion term. The z's are the displacements of the various atoms from their undisturbed coplanar positions. K_1 and K_2 are obtained from the vibrational spectrum of benzene and their values are $0{\cdot}1474\times10^5$ dyn cm^{-1} and $0{\cdot}0551\times10^5$ dyn cm^{-1} respectively. The constant K_2 is probably a good value to use for graphite, but the value of K_1 is less certain, as it involves the out-of-plane motion of hydrogen atoms instead of carbon. However, the difference in the value of K_1 for hydrogen and carbon atoms is not expected to be large.

§ 3. EQUATIONS OF MOTION

The lattice of a single layer of graphite is illustrated in figure 1. The equations of motion for the two atoms in the cell l, m (with coordinates z_{lm}' and z_{lm}'') are

$$M\frac{d^2z_{lm}'}{dt^2}+Q_{lm}'=0, \qquad M\frac{d^2z_{lm}''}{dt^2}+Q_{lm}''=0$$

where M is the mass of a carbon atom, d^2z_{lm}'/dt^2 is the second derivative of the vertical displacement of the single primed atom with respect to time, and Q_{lm}' is the partial derivative of the potential function of the same atom with respect to the same displacement.

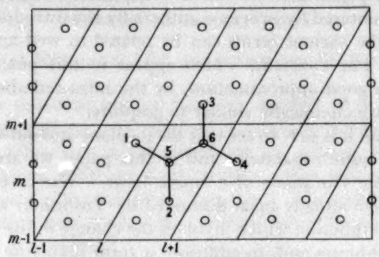


Figure 1. Graphite lattice.

If we use the cyclic boundary conditions of Born with $2N+1$ cells in each of the fundamental directions, the solutions of these equations are of the form

$$Mz_{lm}' = A\exp(i\omega t)\exp\left[2\pi i\left(\frac{xl}{2N+1}+\frac{ym}{2N+1}\right)\right]$$

$$Mz_{lm}'' = B\exp(i\omega t)\exp\left[2\pi i\left(\frac{xl}{2N+1}+\frac{ym}{2N+1}\right)\right]$$

with $l,\ m=-N,\ -(N-1),\dots,0,\dots,N$ and $\omega=2\pi\nu$. Because of the cyclic boundary conditions the values of x and y are then restricted to $x, y=0, 1,\dots N$. Since N is a very large number, the discrete variables can be replaced by continuous ones which vary uniformly from 0 to 2π. Substituting for z' and z'' into the equations of motion leads to the secular equations

$$(-M\omega^2+\alpha)A+\beta B=0, \qquad \bar{\beta}A+(-M\omega^2+\alpha)B=0$$

where α and β are rather complicated functions of the variables x and y. The solution of these two simultaneous equations is given by $(-M\omega^2+\alpha)^2=\beta\bar{\beta}$ with

$$\alpha=9(3K_1+2K_2)+(3K_1-2K_2)G^2(x,y)$$

$$\begin{aligned}\beta\bar{\beta}=81K_1^2G^2(x,y)&+18K_1K_2[G^2(2x,2y)-G^2(x,y)]\\&+K_2^2[27+G^2(3x,3y)+5G^2(2x,2y)-3G^2(x,y)\\&-4G^2(2x-y,x-2y)-G^2(3x-2y,x-3y)-G^2(2x-3y,3x-y)]\end{aligned}$$

and

$$G^2(x,y)=2\cos x+2\cos y+2\cos(x-y)+3.$$

In order to obtain the distribution of vibrations, the secular equations must be solved for ω in terms of the variables x and y, treating the two branches separately. Figures 2 and 3 show plots of the contours of constant ω for various values of the coordinates x and y. The distribution function, which is the number of vibrations contained in an interval of frequency between ω and $\omega+d\omega$, is proportional to the area contained between the two contours ω and $\omega+d\omega$. The constant of proportionality is the normalization factor which makes the total area equal to unity.

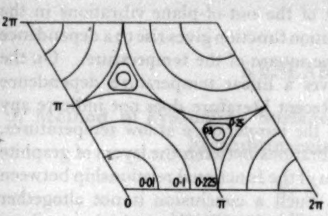


Figure 2. Contours of constant $\omega^2/\omega_{\max}{}^2$ for lower branch.

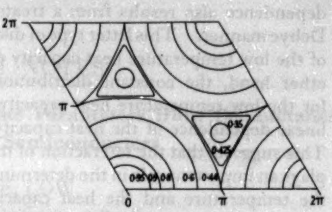


Figure 3. Contours of constant $\omega^2/\omega_{\max}{}^2$ for upper branch.

§ 4. Results and Discussion

The distribution function of the vibrations perpendicular to the plane of a layer of graphite for a Coulson–Senent type of potential function is shown in figure 4. There is a remarkable similarity between this diagram and the corresponding diagram (Coulson and Taylor 1952) for the distribution of energies of π-electrons in a single graphite layer. Formally, of course, the two problems have considerable similarity. This distribution had to be obtained by numerical integration because of the difficulty in solving the secular equations analytically. The general features of the distribution agree with the predictions of van Hove (1953) in that there are two singularities in the distribution corresponding to each of the saddle points in the wave surfaces shown in figures 2 and 3. It is interesting to note that the distribution approaches a constant value for small values of $\nu/\nu_{\max}$. This result is not in agreement with the calculations of Rosenstock (1953), but this is not unexpected because the two models chosen to represent the force field in graphite differ considerably. However, Newell (1955, 1956) has discussed this type of result in relation to a distinction between valence forces and nearest-neighbour forces. In the former case the distribution function is constant for small values of $\nu/\nu_{\max}$ whereas in the latter the distribution approaches zero in a linear fashion for small values of $\nu/\nu_{\max}$. The linear

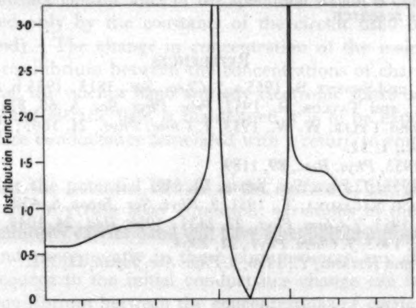


Figure 4. Distribution function for out-of-plane vibrations.

dependence also results from a treatment of the out-of-plane vibrations in the Debye manner. This latter type of distribution function gives rise to a dependence of the low temperature heat capacity on the square of the temperature. On the other hand, the constant distribution gives a linear temperature dependence for the low temperature heat capacity. Recent literature does not indicate any linear dependence of the heat capacity on the temperature at low temperatures. This suggests that the interaction of the vibrations between the layers of graphite plays an important role in the determination of the functional relationship between the temperature and the heat capacity. Such a conclusion is not altogether unexpected, since we know that it is the low-frequency vibrations which contribute most significantly at low temperatures; and it is precisely these low-frequency vibrations which will be most sensitive to the inter-layer forces. It does seem probable, however, that apart from this low-frequency failure, our distribution function is substantially correct. Certainly an analogous situation is already known to exist in the electronic band structure of the π-electrons of one layer and of a complete three-dimensional crystal.

ACKNOWLEDGMENTS

One of us (P. B. S.) would like to express his appreciation to the United States Educational Commission in the United Kingdom for making funds available for this research.

REFERENCES

COULSON, C. A., and SENENT, S., 1955 a, *J. Chem. Soc.*, 1813; 1955 b, *Ibid.*, 1819.
COULSON, C. A., and TAYLOR, R., 1952, *Proc. Phys. Soc.* A, **65,** 815.
DESORBO, W., and TYLER, W. W., 1953, *J. Chem. Phys.*, **21,** 1663; 1956, *Bull. Amer. Phys. Soc.*, **1,** 31.
VAN HOVE, L., 1953, *Phys. Rev.*, **89,** 1189.
KOMATSU, K., 1955, *J. Phys. Soc. Japan*, **10,** 346.
KOMATSU, K., and NAGAMIYA, T., 1951, *J. Phys. Soc. Japan*, **6,** 438.
NEWELL, G. F., 1955, *J. Chem. Phys.*, **23,** 2431; 1956, *Ibid.*, **24,** 1049.
ROSENSTOCK, H., 1953, *J. Chem. Phys.*, **21,** 2064.
YOSHIMORI, A., and KITANO, Y., 1956, *J. Phys. Soc. Japan*, **11,** 352.

RESEARCH NOTES

A Method of Evaluating Surface State Parameters from Conductance Measurements on Semiconductors

By G. G. E. Low
Department of Physics, University of Reading

Communicated by R. W. Ditchburn; *MS. received* 12*th June* 1956, *and in final form* 19*th July* 1956

§ 1. Introduction

If an electric field is applied capacitively to the surface of a crystal of extrinsic germanium, the current flow through the specimen which provides the induced charge is chiefly by means of majority carriers as these will predominate in the conduction processes in the bulk. The resulting alteration in the concentration of majority carriers near the surface gives rise to an initial change in the surface conductance of the specimen which is completed in a time normally limited only by the constants of the circuit used to apply the field (e.g. 10^{-8} second). The change in concentration of the majority carriers also leads to a non-equilibrium between the concentrations of charge carriers in the conduction band, the valence band and any electronic states on the surface and consequently if the electric field is maintained it is to be expected that further changes in surface conductance associated with a return to equilibrium will take place.

Provided that the potential barrier at the surface of the germanium is well removed from the inversion layer condition, variations in the concentration of carriers in the minority carrier band will have a negligible influence on the surface conductance and consequently in these circumstances any changes which are observed subsequent to the initial conductance change can be associated with the return to equilibrium between the concentrations of carriers in the majority carrier band and surface states. Experimentally it has been found that in general the above charge redistribution takes place in two steps which can be attributed to two different types of surface states. For well-etched surfaces, the first step is approximately exponential in character and occupies a time of the order of 100 microseconds, the surface conductance at the completion of this process being known as the quasi-stable conductance value. This decay is considered to be due to the trapping action of 'fast' surface states and it is clear that measurements carried out for the conditions mentioned above under which only majority carriers play a significant part in the conductance changes observed can be used to obtain information concerning the densities and cross sections for majority carrier capture of these surface states. Experiments of this nature have been reported recently (Low 1956). In the course of interpreting the results obtained relationships between various surface parameters were employed for which no derivations were given. The purpose of the present note is to supply these mathematical arguments.

§ 2. Surface State Densities

For the conditions indicated above under which only one type of carrier need be considered, an experimental measurement of conductance over a period of a few hundreds of microseconds subsequent to the application of a field gives a result of the form shown in figure 1 (*a*). Three experimental quantities are obtained: the initial conductance change $\delta g_1 + \delta g_2$, the quasi-stable value δg_2 and the time constant associated with the decay which will be denoted by τ. For the sake of simplicity the following discussion will be carried out in terms of a positive charge induced on a p-type semiconductor. Also it will be assumed throughout that the changes in surface potential resulting from the application of the field are small compared with $\boldsymbol{kT/e}$.

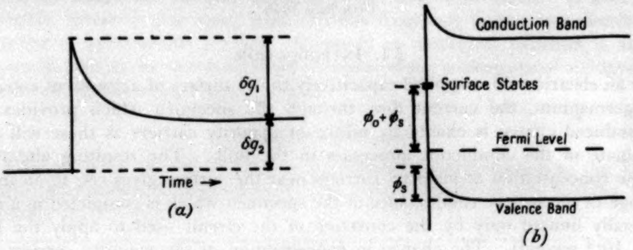


Figure 1.

If the surface density of charge induced on the specimen is δQ_1 then a field effect mobility may be defined operationally by

$$\mu = \frac{\delta g_1 + \delta g_2}{\delta Q_1}. \qquad \ldots\ldots(1)$$

(It is assumed that δg_1 and δg_2 are in units of conductance per square of surface.) This mobility can be used to find the amount of mobile charge present in the valence band which corresponds to a conductance change measured at any subsequent time. Thus, when the quasi-stable state has been reached, the charge remaining untrapped is

$$\delta Q_S = \delta g_2/\mu. \qquad \ldots\ldots(2)$$

This charge can be related to the change in surface barrier height $\delta\phi_S$ by

$$\delta\phi_S \equiv \frac{d\phi_S}{dQ_S}\,\delta Q_S = \frac{d\phi_S}{dQ_S}\,\frac{\delta g_2}{\mu}. \qquad \ldots\ldots(3)$$

(ϕ_S gives the potential for a negative charge, see figure 1 (*b*).)

The charge added to the surface states at any time subsequent to the application of the field is equal to $\boldsymbol{e}\Sigma_i n_i \delta f_{pi}$ where the n_i are the surface densities of various sets of states of different energies and the δf_{pi} are the changes (from their initial values) of the Fermi functions for holes corresponding to these states. Differentiation of the Fermi function shows that for small changes

$$\delta f_{pi} = f_{pi} f_{ni}\,\delta\phi_t \qquad \ldots\ldots(4)$$

where $\delta\phi_t$ is the change in potential of the states relative to their Fermi level in units of $\boldsymbol{kT/e}$. When the quasi-stable state has been reached and the fast surface states are in equilibrium with the valence band again, $\delta\phi_t=\delta\phi_s$. Consequently, the additional charge in the states is given by $\boldsymbol{e}\,\delta\phi_s \sum_i n_i f_{pi} f_{ni}$. This may be equated to the change in free charge in the valence band corresponding to the difference between the initial and the quasi-stable conductance values, i.e.

$$\boldsymbol{e}\,\delta\phi_s \sum_i n_i f_{pi} f_{ni} = \frac{\delta g_1}{\mu}. \qquad \ldots\ldots(5)$$

The use of equation (3) to eliminate $\delta\phi_s$ provides the relationship

$$\boldsymbol{e} \sum_i n_i f_{pi} f_{ni} = \frac{dQ_s}{d\phi_s}\frac{\delta g_1}{\delta g_2} \qquad \ldots\ldots(6)$$

which, by plotting the quantities on the right-hand side against ϕ_s, may be used to obtain information concerning the energies and densities associated with the fast surface states. ($dQ_s/d\phi_s$ can be calculated from ϕ_s so that experimentally it is necessary to measure δg_1 and δg_2 as functions of ϕ_s.)

§ 3. SURFACE STATE CROSS SECTIONS FOR MAJORITY CARRIER CAPTURE

On the assumption that only one set of surface states need be considered, the time constant τ measured experimentally can be related to the capture cross sections of the states as follows (the subscript *i* will be dropped from the equations). The first step in the argument requires that the net rate of capture of holes by the surface states be calculated. By a method strictly analogous to that used by Shockley and Read (1952) when treating recombination in the bulk, it can be shown that this rate is given by

$$U=\frac{n}{\mathrm{T}}(f_n p_s - f_p p_{s1}) \qquad \ldots\ldots(7)$$

where T is a constant whose reciprocal is equal to one-quarter of the product of the cross section of a trap and the thermal velocity of the holes, and where p_s and p_{s1} are the concentrations of holes in the valence band at the surface during the measurement, and when the Fermi level passes through the surface states respectively. For equilibrium

$$f_n p_s = f_p p_{s1} \qquad \ldots\ldots(8)$$

so that for a small displacement

$$\begin{aligned} U &= \frac{n}{\mathrm{T}}\{(f_n+\delta f_n)(p_s+\delta p_s)-(f_p+\delta f_p)p_{s1}\} \\ &= \frac{n}{\mathrm{T}}\{f_n\delta p_s + (p_s+p_{s1})\delta f_n\} \qquad \ldots\ldots(9) \end{aligned}$$

as $\delta f_p = -\delta f_n$. It will be assumed that the quasi-stable state corresponds to the equilibrium condition defined by equation (8), so that at any time before this state is reached the charge in the fast surface states is less than the value for equilibrium by an amount

$$\delta Q_t = \boldsymbol{e}n\,\delta f_n. \qquad \ldots\ldots(10)$$

As the charge in the valence band will be increased by δQ_t equation (9) can be rewritten

$$\boldsymbol{e}U = -\frac{dQ_t}{dt} = \frac{1}{\mathrm{T}}\left(\boldsymbol{e}nf_n\frac{dp_s}{dQ_s}+p_s+p_{s1}\right)\delta Q_t. \qquad \ldots\ldots(11)$$

Thus, the time constant associated with this capture process is given by

$$\frac{1}{\tau}=\frac{1}{\mathrm{T}}\left(enf_{\mathrm{n}}\frac{dp_{\mathrm{s}}}{dQ_{\mathrm{s}}}+p_{\mathrm{s}}+p_{\mathrm{s}1}\right). \qquad \ldots\ldots(12)$$

The quantity n may be eliminated from equation (12) with the aid of equation (6). Further simplifications may be made by the use of the relations

$$p_{\mathrm{s}}=p_0\exp(\phi_{\mathrm{s}}), \quad p_{\mathrm{s}1}=p_0\exp(-\phi_0) \qquad \ldots\ldots(13)$$

where p_0 is the concentration of holes in the valence band in the bulk of the specimen and ϕ_0 is the difference in energy between the states and the Fermi level when $\phi_{\mathrm{s}}=0$ (see figure 1(*b*)). It follows from these that

$$dp_{\mathrm{s}}/d\phi_{\mathrm{s}}=p_{\mathrm{s}}, \quad p_{\mathrm{s}1}/p_{\mathrm{s}}=\exp(-\phi_0-\phi_{\mathrm{s}}) \qquad \ldots\ldots(14)$$

so that equation (12) can finally be reduced to

$$\frac{1}{\tau}=\frac{1}{\mathrm{T}}\left\{\frac{\delta g_1}{\delta g_2}+\frac{1+\exp(-\phi_0-\phi_{\mathrm{s}})}{f_{\mathrm{p}}^{-1}}\right\}p_{\mathrm{s}}f_{\mathrm{p}}^{-1}. \qquad \ldots\ldots(15)$$

Thus, the surface state parameter T can now be written

$$\mathrm{T}=\tau\left\{\frac{\delta g_1}{\delta g_2}+\frac{1+\exp(-\phi_0-\phi_{\mathrm{s}})}{f_{\mathrm{p}}^{-1}}\right\}f_{\mathrm{p}}^{-1}p_0\exp(\phi_{\mathrm{s}}) \qquad \ldots\ldots(16)$$

and the cross section for majority carrier capture may be evaluated.

One point which arises in connection with the use of the above equations concerns the form of the Fermi function appropriate to the fast surface states. It has been pointed out by Statz *et al.* (1956) that the states in question are probably of the type which can accommodate one electron of either spin but not two electrons of opposite spin. In this case their occupancy is given by the function

$$\frac{\frac{1}{2}\exp(\phi_0+\phi_{\mathrm{s}})}{1+\frac{1}{2}\exp(\phi_0+\phi_{\mathrm{s}})}. \qquad \ldots\ldots(17)$$

No matter which Fermi function is appropriate, it is clear that the probability of finding a hole in a state well above the Fermi level will closely approach unity. Thus, in this case equation (16) may be reduced to

$$\mathrm{T}=\tau(\delta g_1/\delta g_2+1)p_0\exp(\phi_{\mathrm{s}}). \qquad \ldots\ldots(18)$$

Further, if the states lie below and again well removed in energy from the Fermi level the probability of occupancy by a hole will have an energy dependence which is given always approximately by the Boltzmann expression $\exp(\phi_0+\phi_{\mathrm{s}})$.

ACKNOWLEDGMENTS

The author wishes to thank Professor R. W. Ditchburn for his support, Dr. P. C. Banbury for a number of valuable discussions, and the New Zealand Defence Research Organisation for financial maintenance.

REFERENCES

LOW G. C. E., 1957, *Report of the Conference on Semiconductors, Rugby* (London: Physical Society), in the press.

SHOCKLEY, W., and READ, W. T., 1952, *Phys. Rev.*, **87**, 835.

STATZ, H., DE MARS, G. A., DAVIS, L., and ADAMS, A., 1956, *Phys. Rev.*, **101**, 1272.

The Statistics of Divalent Impurity Centres in a Semiconductor

By C. H. CHAMPNESS
Research Department, Metropolitan-Vickers Electrical Co., Ltd., Manchester

MS. received 26*th March* 1956 *and in revised form* 2*nd August* 1956

§ 1. INTRODUCTION

In a semiconductor in which the impurities at absolute zero of temperature have each two electrons (or holes) above that required for bonding, two activation energies to the conduction band arise corresponding to two levels below the conduction band. For two such levels the statistical treatment is formally different from that for two independent levels. This is because in the case of a doubly charged centre the number of states in one level is neither constant nor independent of the number of states in the other level and depends on whether the centre is neutral, singly ionized or doubly ionized. However, the total number of states altogether from both levels is constant and equal to twice the number of impurities. The problem can be treated in a manner analogous to that for singly charged centres by counting the states in such a way as to take into account the double degeneracy due to spin and applying the normal Fermi distribution function. Alternatively, a simpler way of counting the states can be used with occupation probabilities which already take account of spin degeneracy. A third method is the more fundamental one of finding the condition for minimum free energy of the electrons. All three methods will be found to give the same equation for the concentration of electrons in the conduction band as a function of temperature and impurity concentration. However only the first method is treated in detail here as it shows greater physical insight into the problem. Because the two levels considered are well separated from each other, the final results are not materially different from what would be obtained by considering two independent levels (caused by identical concentrations of two different impurities).

§ 2. DERIVATION OF EQUATION FOR ELECTRON CONCENTRATION

We consider two levels. The upper one (that nearest to the conduction band) represents the energy of each of the two electrons on a neutral centre and the lower level the energy of the single electron on a singly ionized centre.

N is the density of impurity centres, M_0, M_1, M_2 that of neutral, singly ionized and doubly ionized impurity centres respectively, n that of electrons in the conduction band and n_1, n_2 that of electrons in the upper and lower levels respectively. f_1, f_2 are the Fermi functions at the upper and lower levels and N_1, N_2 are the density of states in the upper and lower levels.

The statistics can be treated by the following method of counting states using the normal Fermi distribution function. A neutral centre carries two electrons of opposite spin, either of which can jump from the upper level to the conduction band (the other falling to the lower level) and may therefore be counted as providing two filled states at the upper level. A singly ionized centre can lose one electron from the lower level to the conduction band or it can accept another electron of opposite spin thus raising the energy of both electrons to the upper level. A singly ionized centre may therefore be counted as providing one empty state to the upper level and at the same time one filled state to the lower level. A doubly ionized centre can accept an electron of either spin which will go into the lower level and is therefore counted as providing two empty states at the lower level. Thus at the upper and lower levels we have respectively

$$\left.\begin{aligned} N_1 &= 2M_0 + M_1 \quad \text{and} \quad n_1 = 2M_0 \\ N_2 &= M_1 + 2M_0 \quad \text{and} \quad n_2 = M_1. \end{aligned}\right\} \qquad \ldots\ldots(1)$$

The total number of states on both levels is thus

$$2(M_0 + M_1 + M_2) = 2N.$$

The Fermi function at the upper level is therefore

$$f_1 = \frac{n_1}{N_1} = \frac{2M_0}{2M_0 + M_1} \qquad \ldots\ldots(2)$$

and the Fermi function at the lower level is

$$f_2 = \frac{n_2}{N_2} = \frac{M_1}{M_1 + 2M_2}\,. \qquad \ldots\ldots(3)$$

(The above equations give $N_2 f_2 = N_1(1 - f_1)$, i.e. number of electrons on the lower level = number of holes on the upper level.) Now the total number of electrons in the conduction band is given by

$$n = 2M_2 + M_1 = M_1/f_2 \qquad \ldots\ldots(4)$$

and also by

$$n = \int_{E_0}^{\infty} 2D(E) f(E)\, dE \qquad \ldots\ldots(5)$$

where $D(E)$ is the density of states in the conduction band and $f(E)$ is the Fermi function.

Combining equations (2), (3), (4) and (5) with the fact that $N = M_0 + M_1 + M_2$, the number of electrons is given by

$$n = \frac{2N(1 - f_1)}{f_2 - f_1 + 1} = \int_{E_0}^{\infty} 2D(E) f(E)\, dE \qquad \ldots\ldots(6)$$

and in this equation

$$f_1 = \left[\exp\left(\frac{E_1 - \zeta}{kT}\right) + 1\right]^{-1}, f_2 = \left[\exp\left(\frac{E_2 - \zeta}{kT}\right) + 1\right]^{-1}, f(E) = \left[\exp\left(\frac{E - \zeta}{kT}\right) + 1\right]^{-1} \qquad \ldots\ldots(7)$$

where E_0, E_1 and E_2 are the energies of the bottom of the conduction band, the upper and lower levels respectively and ζ is the Fermi level, all measured from an arbitrary energy zero.

Equation (6) may also be obtained by counting the states in a different way from relations (1) and by using the occupation probabilities given by Wilson (1953) viz.

$$f_1' = \left[2\exp\left(\frac{E_1 - \zeta}{kT}\right) + 1\right]^{-1} \quad \text{and} \quad f_2' = \left[\tfrac{1}{2}\exp\left(\frac{E_2 - \zeta}{kT}\right) + 1\right]^{-1}$$

Wilson 1953, Appendix A.1, cases (iii) and (ii)).

At the first level we then count the number of states as $N_1' = M_0 + M_1$ and the number of electrons on them as $n_1' = M_0 = N_1 f_1'$. At the second level we count the states as $N_2' = M_1 + M_2$ and the number of electrons as $n_2' = M_1$ (as in the first method). Using these relations instead of relations (1) and corresponding changes in the other equations with f_1 changed to f_1' and f_2 to f_2' we obtain the result

$$n = \frac{N(2-f_2')(1-f_1')}{1-f_1'+f_1'f_2'} = \int_{E_0}^{\infty} 2D(E)f(E)\,dE.$$

This reduces to equation (6) on substitution for f_1' and f_2'.

An alternative treatment is to evaluate and minimize the free energy of the electrons. This has been done generally by Landsberg (1952) for independent levels and has been applied to this case by R. Stratton (private communication) again yielding equation (6).

§ 3. Solution of Equation (6)

Equation (6) with substitution from equations (9) has been solved numerically for the case $E_0 - E_2 = 0{\cdot}2$ ev, $E_0 - E_1 = 0{\cdot}01$ ev and the ratio of the conduction to the free electron mass $m^*/m = 1$. The results are shown in figure 1. Here the relative electron concentration for the conduction band is plotted against reciprocal absolute temperature for various impurity concentrations.

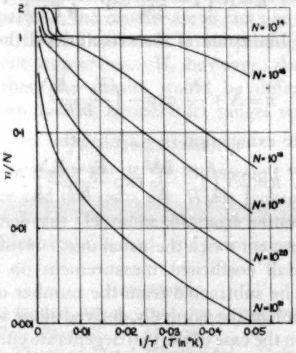


Figure 1. Variation of the degree of ionization with temperature for a semiconductor with helium-like levels at 0·01 ev and 0·2 ev below the conduction band in which $m^*/m = 1$.

Under non-degenerate conditions (i.e. when $E_0 - \zeta > \boldsymbol{k}T$) and when the two levels are well separated equation (6) reduces to the following cubic:

$$x^3 + 2\alpha x^2 - 2\alpha(1-\beta)x - \alpha\beta = 0$$

where

$$x = \frac{n}{2N} = \frac{N_e}{2N}\exp\left(-\frac{E_0-\zeta}{\boldsymbol{k}T}\right), \quad \alpha = \frac{N_e}{2N}\exp\left(\frac{E_1-E_0}{\boldsymbol{k}T}\right)$$

$$\beta = \frac{N_e}{2N}\exp\left(\frac{E_2-E_0}{\boldsymbol{k}T}\right) \quad \text{and} \quad N_e = 2\left(\frac{2\pi m^* \boldsymbol{k}T}{h^2}\right)^{3/2}.$$

This equation is readily soluble in special cases. At lower temperatures in which $n \leqslant N$ we get

$$\frac{n}{N}=\frac{2}{1+(1+1/\alpha)^{1/2}} \qquad \ldots\ldots(8)$$

and at higher temperatures in which $n \geqslant N$ we get

$$\frac{n}{N}=\frac{4\beta}{(1-\beta)[\pm\{1+8\beta/(1-\beta)^2\}^{1/2}-1]}. \qquad \ldots\ldots(9)$$

These solutions apply to the upper three curves of figure 1 ($N=10^{14}$, 10^{16} and 10^{18} centres cm^{-3}) which are non-degenerate. The rest of the curves are for the most part degenerate.

§ 4. DISCUSSION

The two expressions (8) and (9) enable the variation of n with temperature to be traced. At low temperatures ionization of the first level takes place according to the equation

$$n=\sqrt{2}N^{1/2}N_e{}^{1/2}\exp\left(-\frac{E_0-E_1}{2\boldsymbol{k}T}\right) \qquad \ldots\ldots(10)$$

(differing by a factor $\sqrt{2}$ from the usual expression for hydrogen-like impurities) and approaches completion with

$$n=N\left(1-\frac{N}{2N_e}\exp\frac{E_0-E_1}{\boldsymbol{k}T}\right)$$

until $n=N$ is reached. Ionization of the second level then begins according to the equation

$$n=N+\frac{N_e}{2}\exp-\left(\frac{E_0-E_2}{\boldsymbol{k}T}\right) \qquad \ldots\ldots(11)$$

and approaches complete exhaustion ($n=2N$) with

$$n=2N\left(1-\frac{2N}{N_e}\exp\frac{E_0-E_2}{\boldsymbol{k}T}\right).$$

It should be especially noted from equation (11) that ionization from the lower level occurs with an exponent twice the usual one. Hence to obtain the second ionization level from Hall coefficient measurement on samples with divalent impurities N must first be subtracted from the number of carriers given by the Hall coefficient and then half the normally derived slope activation energy taken.

It will be noted that in the case of the two degenerate curves ($N=10^{20}$ and 10^{21}), two regions corresponding to ionization of the first and then second level do not exist. Hence slopes at high and low temperatures are in such cases not directly meaningful.

These considerations throw doubt on the conclusions drawn (second activation energies in particular) by Breckenridge and Hosler (1953) from their measurements on partially reduced rutile specimens which contain two ground 'impurity' levels. (These authors incorrectly ascribe three ground levels to the 'impurities'.)

Figure 2 where $\log n/N$ is plotted against N, shows that it becomes progressively more difficult to ionize electrons from impurities as the concentration of impurities is increased. At lower temperatures and for $N>10^{18}$ where the classical equation for semiconductors very nearly applies, n/N decreases roughly as $N^{-1/2}$. At higher temperatures, the variation is not so great because of approaching exhaustion (i.e. $n=N$ and $n=2N$).

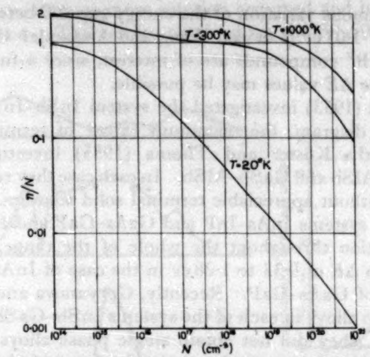


Figure 2. Variation of the degree of ionization with impurity concentration.

Because the two dependent levels arising out of the same impurity considered in this note are well separated, the results are in fact not materially different from conditions arising from two independent levels produced by identical concentrations of two different impurities. If, however, the two dependent levels were close together, then the results would be different. In particular, in non-degenerate cases two distinct temperature ranges would not be present.

ACKNOWLEDGMENTS

The author wishes to thank Dr. R. W. Sillars and Dr. R. Stratton for their interest, valuable advice and criticism, Dr. Willis Jackson, Director of Research and Education and Mr. B. G. Churcher, Manager of the Research Department, Metropolitan-Vickers Electrical Co. Ltd., for permission to publish this research note.

REFERENCES

BRECKENRIDGE, R. G., and HOSLER, W. R., 1953, *Phys. Rev.*, **91**, 793.
LANDSBERG, P. T., 1952, *Proc. Phys. Soc.* A, **65**, 604.
WILSON, A. H., 1953, *The Theory of Metals* (**Cambridge: University Press**).

Solid Solution in the GaSb-InSb System

BY J. C. WOOLLEY, B. A. SMITH AND D. G. LEES
Department of Physics, University of Nottingham

Communicated by L. F. Bates; *MS. received* 3rd *July* 1956

A CONSIDERABLE amount of work has been carried out recently on crystal systems composed of two semiconducting materials of the type $A^{III}B^{V}$ such as InSb, GaAs, GaP, etc. In all cases it appears possible to treat such mixtures as quasi-binary systems. The original work of this kind on semiconductors is that of Stöhr and Klemm (1939), who investigated the phase diagram of the binary system Ge–Si and showed that complete solid solution was obtained throughout the whole range of alloys. It has since been shown that corresponding to the continuous range of solid solution in this system there occurs a continuous variation of the energy gap ΔE between those of pure Ge and Si, i.e. 0·69–1·03 ev (Levitas, Wang and Alexander 1954). As a result, mixtures of the $A^{III}B^{V}$ compounds are of interest, since a further extension of the range of available ΔE values may be possible.

Shih and Peretti (1953) investigated the system InSb–InAs and obtained a degenerate eutectic diagram, the maximum extent of terminal solid solution being 2%. Similarly, Köster and Thoma (1955) investigated the systems InSb–GaSb, InSb–AlSb and GaSb–AlSb. In each case they reported degenerate eutectic diagrams without appreciable terminal solid solution. Folberth (1955) has investigated the systems InAs–InP and GaAs–GaP and for each has found complete solid solution throughout the whole of the range. This results in ranges of energy gap ΔE of 0·33 to 1·28 ev in the case of InAs–InP and 1·45 to 2·25 ev in the case of GaAs–GaP. Recently, Goryunova and Fedorova (1955) have investigated two alloys in each of the systems InSb–GaSb and InAs–GaAs. In the latter system, they did not obtain single phase alloys with the time of anneal used, but did notice some change in the form of the x-ray diffraction picture for the alloy as annealing proceeded. In the case of InSb–GaSb, in contrast with the results of Köster and Thoma, Goryunova and Fedorova found that the two alloys investigated became single phase after a considerable time of anneal at high temperature.

Various suggestions have been put forward to give a general rule fitting these results; Baruch and Desse (1955) suggest that compounds having the group III element common, e.g. InAs–InP, form a continuous range of solid solution, while compounds with the group V element common, e.g. InSb–GaSb, are immiscible in the solid phase. Such rules, however, never cover all the available experimental results.

The work on the system GaSb–InSb described here has been carried out as a preliminary step in the more complete investigation of the crystallographic and electrical properties of the various systems mentioned above. This system was chosen for the first investigation because of the relative ease with which the component compounds can be obtained.

Alloys containing 10, 20, 40, 50, 60, 80 and 90 molecular per cent GaSb were produced by weighing the appropriate amounts of GaSb and InSb, sealing under vacuum in quartz, heating to approximately 1000°C, quenching in water, and giving a preliminary short anneal. The resulting ingots were then powdered and x-rayed using Cu Kα radiation. In every case it was found that the specimens were closer to the two-phase rather than the single-phase state, but that the x-ray photographs were not characteristic of a material consisting of two completely separate phases. Instead of sharp lines, broad bands were obtained, the higher and lower angle edges of the band in any particular case corresponding approximately to the positions of the equivalent lines in the photographs for GaSb and InSb respectively. The intensity distribution across such a band as measured by microphotometer is shown in figure 1 (*a*). It would appear that a range of compositions is present, the two peaks in the distribution curve corresponding to GaSb with a little InSb in solution and InSb containing a little GaSb.

The alloys were then annealed in the powder form following the technique given by Stöhr and Klemm, i.e. heating first at a temperature just below the melting temperature of the lower melting point component, and then slowly increasing the temperature as more solution occurred and hence the lower limit for melting increased. The results obtained on the 50–50 alloy are typical of the whole range. The alloy was annealed successively at 480°c, 520°c and 540°c for varying intervals of time. The intensity distribution of the 422 line plotted as a function of Bragg angle is shown in figure 1 for photographs taken after stated times of anneal. It is seen that the original band slowly narrows and that after some 1300 hours anneal the material is practically single phase.

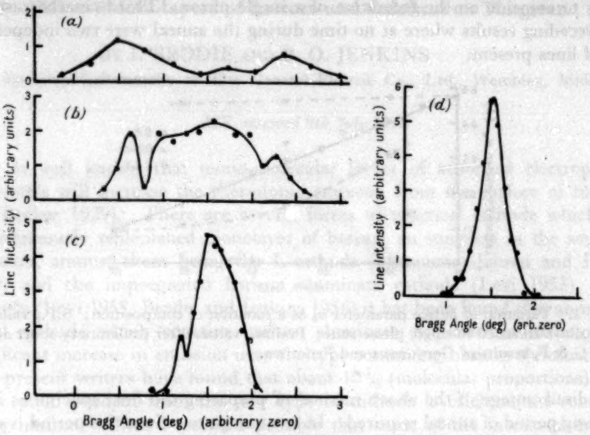


Figure 1. Profile of 422 line for 50% GaSb–50% InSb alloy after various periods of annealing; (*a*) after 18 hours at 480°c, (*b*) after 214 hours at 480°c and 120 hours at 520°c, (*c*) after 214 hours at 480°c, 170 hours at 520°c and 260 hours at 540°c, (*d*) after 214 hours at 480°c, 170 hours at 520°c and 910 hours at 540°c.

Similar results were obtained with all the other alloys in this system. Each alloy was brought to the single phase state by appropriate annealing and then investigated by x-ray methods. In every case the zinc-blende structure was found to be retained, and the cell parameters a_0 were calculated. A graph of a_0 against molecular percentage of GaSb, including the results for pure GaSb and InSb, is shown in figure 2. It is seen that within the limits of experimental error there is a linear variation of lattice parameter with composition.

In the case of the alloys containing 20, 50 and 80% GaSb, x-ray measurements were made after the short preliminary anneal, and measurements made to find the effective values of lattice parameters corresponding to the high and low angle edges of the bands in each case. The values are plotted in figure 2. It is seen that the values of the parameter at the InSb edge of the band vary very little with composition, while those for the GaSb edge increase with increased InSb content. This would appear to indicate a more rapid diffusion of In into the GaSb lattice than for Ga into the InSb lattice.

One point of comparison with the Ge–Si system is the form of the x-ray photographs during the non-equilibrium stage of the anneal. It would appear from the description by Stöhr and Klemm that in the case of Ge–Si, x-ray lines were obtained corresponding to a two-phase system rather than to the band form described above, and that these lines moved together as the annealing proceeded, finally to coincide on the formation of a single phase. This is in contrast with the preceding results where at no time during the anneal were two independent sets of lines present.

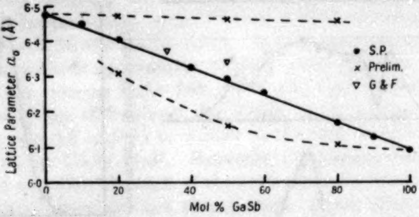


Figure 2. Variation of lattice parameter a_0 as a function of composition; S.P., values for alloys annealed to single phase state; Prelim., values after preliminary short anneal; G & F, results of Goryunova and Fedorova.

A disadvantage of the above method of preparing the final specimens is the very long period of anneal required. In an attempt to reduce this period, various methods of preparing the initial alloys were tried. These included chill casting in a water-cooled copper mould, slow cooling from just below the solidus, and very slow cooling (approximately 1 degree per hour) through the liquidus–solidus range, but in no case was there any appreciable change in the form of the unannealed material. The annealing was also carried out with the specimen as a solid ingot, as a powder and as powder compressed into capsules. The last of these methods seemed to give some reduction in the required time of anneal, but even in this case the improvement was only small.

It would appear from the present results that solid solution occurs throughout the whole range of the GaSb–InSb system and that the apparent eutectic form obtained by Köster and Thoma was due to insufficient annealing of the alloys concerned. Further work is being carried out to determine the complete phase diagram of the GaSb–InSb system and also to determine the variation of energy gap as a function of composition. Similar work is also being carried out on other mixed $A^{III}B^{V}$ systems and it is hoped to publish details later.

ACKNOWLEDGMENTS

Grateful acknowledgment is made to Professor L. F. Bates for the facilities of his laboratory, to Mr. J. T. Edmond for helpful discussions, and to the Chief Scientist, Admiralty, for sponsoring the work.

REFERENCES

BARUCH, P., and DESSE, M., 1955, *C. R. Acad. Sci., Paris*, **241**, 1040.
FOLBERTH, O. G., 1955, *Z. Naturf.*, **10a**, 502.
GORYUNOVA, N. A., and FEDOROVA, N. N., 1955, *J. Tech. Phys., Moscow*, **24**, 1339.

Köster, W., and Thoma, B., 1955, *Z. Metallk.*, **46**, 293.
Levitas, A., Wang, C. C., and Alexander, B. H., 1954, *Phys. Rev.*, **95**, 846.
Shih, C., and Peretti, E. A., 1953, *J. Amer. Chem. Soc.*, **74**, 608.
Stöhr, H., and Klemm, W., 1939, *Z. anorg. Chem.*, **241**, 305.

Mixed Monolayers of Barium and Calcium on Tungsten

By I. BRODIE and R. O. JENKINS
Research Laboratories of The General Electric Co., Ltd., Wembley, Middx.

MS. received 9th July 1956

It is well known that mono-molecular layers of adsorbed electropositive metals will increase the thermionic emission from the surface of tungsten (Becker 1929). There are several forms of practical cathode which have a continuously replenished monolayer of barium on tungsten as the source of emission, amonst them being the L-cathode (Lemmens, Jansen and Loosjes 1950) and the impregnated barium aluminate cathode (Levi 1953). More recently (Levi 1955, Brodie and Jenkins 1956) it has been found that an impregnated cathode which contains barium aluminate and calcium oxide gives a significant increase in emission over the type containing barium aluminate only. The present writers have found that about 10% (molecular proportions) of the evaporation product from the new type of cathode is calcium, the remainder being barium. Thus it seemed possible that some of the increased emission obtained from these cathodes might be due at least in part to a mixed monolayer of barium and calcium on tungsten.

Measurements on mixed monolayers have, therefore, been made. The procedure was as follows. A mixed vapour of barium and calcium was condensed on a tungsten wire which was held at a constant temperature. The saturated thermionic emission from this wire was measured as a function of the time the wire was exposed to the mixed vapour. The usual form of Becker curve (Becker 1929) in which the emission passes through a maximum when the monolayer is complete, was obtained provided that the temperature of the wire did not exceed 750°K. The non-existence of a clear maximum at higher wire temperatures was due to migration of the calcium which begins at about 750°K, barium migration only becoming significant at about 900°K.

Two methods were used to estimate the fraction of the monolayer which was calcium. In the first method the barium evaporation rate was kept constant throughout a run so that the time for the barium alone to reach a monolayer was known. The difference between this time and the actual time a mixed vapour took to reach the monolayer condition gave the fraction of the layer which was calcium. In the second method, which gave good agreement with the first, the times for the calcium and barium to reach the monolayer condition were measured separately before operating the sources simultaneously. The source of barium was a commercial getter consisting of a mixture of barium titanate and beryllium applied to a molybdenum strip, and the source of calcium was calcium oxide applied to a tantalum strip. The metal strips could be heated separately so that the rate of chemical reactions which produced the free barium and calcium could be controlled.

In figure 1 the value of the maximum (complete monolayer) emission obtained from the wire is plotted against the fraction of the layer that is calcium. The maximum emission is obtained when about a quarter of the monolayer is calcium, and is more than double that obtained from a barium monolayer and almost six times that from a calcium monolayer. If it is assumed that the changes in emission are due to changes in the work function of the surface only, then for the wire held at a given temperature T

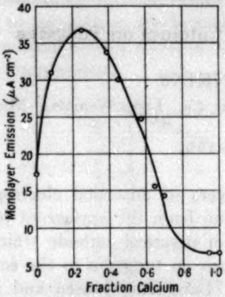


Figure 1. Monolayer emission from a mixed layer of barium and calcium on tungsten at 750°K.

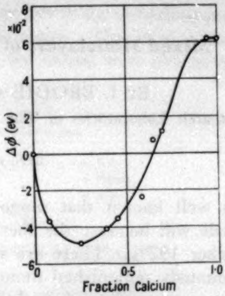


Figure 2. Work function of a mixed monolayer of barium and calcium on tungsten.

$$\Delta\phi = -\boldsymbol{k}T\ln(I_{\mathrm{mixture}}/I_{\mathrm{Ba}})$$

where I_{mixture} is the value of monolayer emission from a mixture and I_{Ba} that from barium only. Figure 2 shows $\Delta\phi$ calculated in this way plotted against the fraction of calcium in the monolayer.

There are several possible explanations for the increase in emission obtained from the optimum mixed monolayer. It may be that the calcium is preferentially adsorbed on those crystal planes on which barium is not easily adsorbed, or it may be that an actual mixed monolayer on certain crystal faces gives the increased emission. A powerful method for resolving this point would be the field emission microscope (Muller 1938, Benjamin and Jenkins 1942).

Further investigations are being carried out on mixed monolayers of barium and strontium, and of strontium and calcium.

ACKNOWLEDGMENT

Acknowledgment is made to the Director of Physical Research, Admiralty, for permission to publish this account.

REFERENCES

BECKER, J. A., 1929, *Trans. Amer. Electrochem. Soc.*, **55,** 129.
BENJAMIN, M., and JENKINS, R. O., 1942, *Proc. Roy. Soc.* A, **180,** 225.
BRODIE, I., and JENKINS, R. O., 1956, *J. Appl. Phys.*, **27,** 417.
LEMMENS, H. J., JANSEN, M. J., and LOOSJES, R., 1950, *Philips Tech. Rev.*, **11,** 249.
LEVI, R., 1953, *J. Appl. Phys.*, **24,** 233; 1955, *Ibid.*, **26,** 639.
MULLER, E. W., 1938, *Z. Phys.*, **108,** 668.

On the Bose–Einstein Liquid Model for ^{3}He–^{4}He Solutions

By B. K. AGARWAL†

Physics Department, University of Allahabad, India

MS. received 17*th April* 1956

In a previous note (Agarwal 1956) it was shown that the form of the energy spectrum, when modified to include the effect of distortion by the liquid state of the ^{3}He–^{4}He mixture, leads to improved agreement with experimental data on the variation of lambda temperature with ^{3}He concentration. The generalized energy spectrum employed for this was $E=Ap^{1/r}$. It was found that if r is taken to be a function of ^{3}He concentration, the theoretical curve is in good agreement with the observations of Abraham, Weinstock and Osborne (1949) and of Daunt and Heer (1950), on the shift of lambda point due to ^{3}He admixture. However, recent experiments by King and Fairbank (1953), Dokoupil *et al.* (1954) and Dash and Taylor (1955) have shown that earlier observations were largely in error. Therefore, the choice of the function r needs revision. It is seen that a function of the type

$$r=r_0V_3^0/V_4^0, \qquad \ldots\ldots(1)$$

where r_0 is the value of r for perfect gas spectrum, and V_3^0 and V_4^0 are the atomic volumes per atom of ^{3}He and ^{4}He respectively, yields a slope of the lambda temperature against ^{3}He concentration curve in very good agreement with the recent measurements. The function r, defined by equation (1), is independent of ^{3}He concentration. Taking $NV_3^0=37{\cdot}6\ \mathrm{cm}^3$ and $NV_4^0=27{\cdot}6\ \mathrm{cm}^3$ ($N=$ Avogadro's number) for the molar volumes of respective isotopes, we obtain for the theoretical shift of lambda temperature, at very low concentrations, a value of $-1{\cdot}46$ degrees per mole. This may be favourably compared with the following experimental values: $-1{\cdot}47\pm0{\cdot}03$ degrees per mole due to Dash and Taylor (1955), $-1{\cdot}48$ degrees per mole due to Dokoupil *et al.* (1954) and $-1{\cdot}50$ degrees per mole due to King and Fairbank (1953). The table shows the values of lambda temperature of mixture T_λ calculated according to the present model for three different concentrations together with those observed by Dash and Taylor (1955). The agreement of the calculated values with those observed is almost within the experimental error.

Table 1. Lambda Temperature of Mixture

X(%)	Dash & Taylor Exp. (°K)	Calculated (°K)
$3{\cdot}65\pm0{\cdot}2$	$2{\cdot}127\pm0{\cdot}002$	2·129
$6{\cdot}00\pm0{\cdot}2$	$2{\cdot}096\pm0{\cdot}002$	2·095
$9{\cdot}20\pm0{\cdot}2$	$2{\cdot}045\pm0{\cdot}002$	2·048

For a further test of equation (1) the values of total vapour pressure of mixture p ($=p_3+p_4$) have been computed for two different strengths of solution of liquid ^{3}He in ^{4}He. According to the present model the expressions for partial vapour pressures p_3 and p_4 of ^{3}He and ^{4}He respectively can be written as:

---

†Now at the Optics Institute, University of Rochester, N.Y.

$T<T_\lambda$

$$p_3=p_3^0C_3\exp\left[C_4-\frac{\zeta(3r+1)}{\zeta(3r)}\left(\frac{V_3^0}{V_4^0}\right)\left(\frac{T}{T_\lambda}\right)^{3r}C_4\right], \qquad \ldots\ldots(2)$$

$$p_4=p_4^0\exp\left[-\frac{V_4^0}{V_3^0}C_3\right], \qquad \ldots\ldots(3)$$

$T>T_\lambda$

$$p_3=p_3^0C_3\exp\left[\left(1-\frac{V_3^0}{V_4^0}\right)C_4+b_2\zeta(3r)\frac{V_3^0}{V_4^0}\left(\frac{T_\lambda}{T}\right)^{3r}C_4 + b_3\{\zeta(3r)\}^2\frac{V_3^0}{V_4^0}\left(\frac{T_\lambda}{T}\right)^{6r}C_4+\ldots\right], \qquad \ldots\ldots(4)$$

$$p_4=p_4^0C_4\exp\left[\left(1-\frac{V_4^0}{V_3^0}\right)C_3-b_2\zeta(3r)\left\{1+C_3-\frac{1}{C_4}\right\}\left(\frac{T_\lambda}{T}\right)^{3r} -\tfrac{1}{2}b_3\{\zeta(3r)\}^2\left\{1+2C_3-\frac{1}{(C_4)^2}\right\}\left(\frac{T_\lambda}{T}\right)^{6r}-\ldots\right], \qquad \ldots\ldots(5)$$

where $C_3=N_3V_3^0/(N_3V_3^0+N_4V_4^0)$, $C_4=N_4V_4^0/(N_3V_3^0+N_4V_4^0)$, ζ is the Riemann zeta-function, p_3^0 and p_4^0 are the full vapour pressures of pure liquid ^{3}He and ^{4}He respectively, and b_2, b_3 are given, following Kothari and Singh (1941), by

$$b_2=\frac{1}{2^{3r+1}} \quad \text{and} \quad b_3=\left(\frac{2}{3^{3r+1}}-\frac{1}{4^{3r}}\right).$$

The details of the derivation of above formulae are not included here as they easily follow from the original paper of Heer and Daunt (1951).

In figure 1 we have plotted calculated values of total vapour pressure, $p=p_3+p_4$, against temperature for a 9·49% solution. The experimental points included in the figure are Sommers' (1952) unsmoothed data. The curve computed from the theory of Heer and Daunt (1951) has been also plotted for comparison. It is seen that for low temperatures the modified Bose–Einstein liquid model and Heer and Daunt's theory both give values close to Sommers' experimental points, while at higher temperatures the experimental points collect on the curve due to the modified Bose–Einstein liquid model.

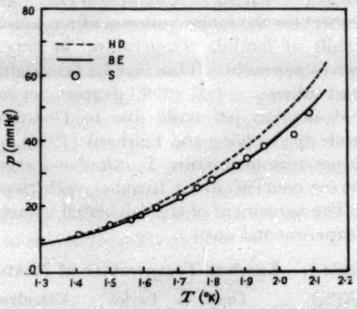


Figure 1. **Relation between total vapour pressure and temperature for a 9·49% solution. HD, Heer and Daunt's theory; BE, modified Bose–Einstein liquid model; S, Sommers' unsmoothed data.**

Figure 2 shows the calculated values of total vapour pressure for a 20% mixture as a function of temperature (full curve) together with the curve derived from the unmodified theory of Heer and Daunt (dotted curve). The experimental points of Weinstock, Osborne and Abraham (1950) for a 20·3% mixture are indicated by circles. As the accuracy of experimental points below 1·6°K is rather limited, it may be noted that for T greater than 1·6°K the theoretical values agree very closely with the observed pressures.

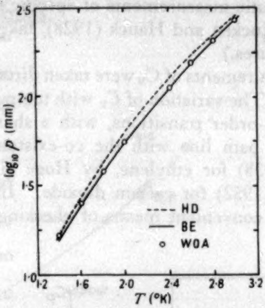


Figure 2. The graph of $\log_{10}p$ against temperature for a 20% mixture of ^{3}He in ^{4}He. HD, Heer and Daunt's theory; BE, modified Bose–Einstein liquid model; WOA, Weinstock, Osborne and Abraham.

Above 1·6°K, the instrumental error in the measurement of vapour pressure is only about ±0·4%. This error is very small in comparison with the difference in the curves according to the present theory and the theory of Heer and Daunt (1951).

In conclusion it may be said that the equation (1) is valid over a wide range of temperature and of ^{3}He concentration.

According to the present model other thermodynamic as well as transport properties of mixtures have been investigated. Theoretical results are in excellent agreement with the experimental data. A detailed paper is to be published elsewhere.

ACKNOWLEDGMENT

The author thanks Dr. G. B. Deodhar for his interest in this work and Dr. Z. Mikura for the valuable correspondence.

REFERENCES

ABRAHAM, B. M., WEINSTOCK, B., and OSBORNE, D. V., 1949, *Phys. Rev.*, **76**, 864.
AGARWAL, B. K., 1956, *Proc. Phys. Soc.* B, **69**, 408.
DASH, J. G., and TAYLOR, R. D., 1955, *Phys. Rev.*, **99**, 598.
DAUNT, J. G., and HEER, C. V., 1950, *Phys. Rev.*, **79**, 46.
DOKOUPIL, Z., VAN SOEST, G., WANSINK, D. H. N., and KAPADNIS, D. G., 1954, *Physica*, **20**, 1181.
HEER, C. V., and DAUNT, J. G., 1951, *Phys. Rev.*, **81**, 447.
KING, J. C., and FAIRBANK, H. A., 1953, *Phys. Rev.*, **91**, 489.
KOTHARI, D. S., and SINGH, B. N., 1941, *Proc. Roy. Soc.* A, **178**, 135.
SOMMERS, H. S., 1952, *Phys. Rev.*, **88**, 113.
WEINSTOCK, B., OSBORNE, D. V., and ABRAHAM, B. M., 1950, *Phys. Rev.*, **77**, 400.

LETTERS TO THE EDITOR

Specific Heats of Fluid Argon near the Critical Point

In the course of an investigation into the (p, V, T) relationships and specific heats of fluid argon near the critical point we have taken measurements of C_V and C_P along the paths indicated by the vertical and horizontal lines of figure 1. (The only other systematic measurements of specific heat of fluid argon appear to have been those of Eucken and Hauck (1928), taken over a limited range of temperatures and pressures.)

As will be seen, measurements of C_V were taken directly across the co-existence curve in certain cases. The variation of C_V with temperature then took the form characteristic of second-order transitions, with a sharp fall in specific heat at the intersection of the path line with the co-existence curve, as also shown by Pall and Maass (1938) for ethylene, by Hoge (1950) for oxygen and by Michels and Strijland (1952) for carbon dioxide. In this case the Ehrenfest relationship provided a convenient means of checking the internal consistency of our data.

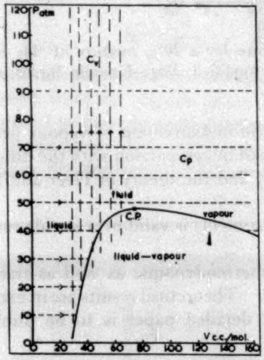


Figure 1.

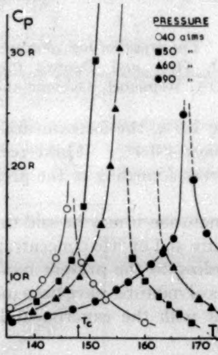


Figure 2.

No previous direct measurements of C_P appear to have been reported for any substance near its critical point. The present results show sharp peaks when C_P is plotted against temperature, the height of a particular peak being very great—corresponding to several hundred R—just above the critical pressure, and decreasing at higher pressures. In figure 2 four such curves are shown, of which one (taken at 40 atmospheres) corresponds to a pressure below the critical value. In this case C_P of course becomes infinite at the point of intersection of the two portions corresponding to the liquid and vapour phases. The other curves shown refer to pressures above the critical value where C_P presumably does not become infinite. Even at 90 atmospheres, however, nearly twice the critical pressure, C_P increases quite sharply by a factor of about five, to a maximum of at least $25R$.

The relationship between the temperature of a given maximum and the pressure of the experiment is of interest. As can be seen in figure 2, these maxima can be located with fair precision, and as far as can be determined within our experimental accuracy, corresponding values of pressure and temperature thus obtained lie exactly on the extrapolation of the vaporization curve on the (p, T) diagram (see figure 3).

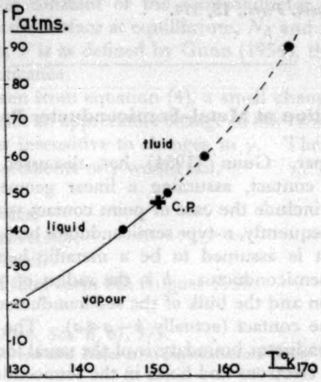


Figure 3.

These data are relevant to the question of the nature of the critical phenomena as discussed, for example, by Mayer and Harrison (1938) and by Rice (1947). The value of a direct measurement of C_p is that it is sensitive, so to speak, right across the top of the co-existence curve. It is of interest to note, as pointed out to us by Dr. R. O. Davies, that the van der Waals equation predicts a maximum of about $5R$ in the curve of $C_p - C_v$ against temperature if C_v is assumed constant, for a pressure of twice the critical value. The calculations of Habgood and Schneider (1954), and of Michels *et al.* (1956), based on their respective (p, V, T) data for xenon, also imply the existence of such maxima as do the data of Hausen (1926) on the Joule–Kelvin coefficients of air. In terms of the (p, T) phase-equilibrium diagram of figure 3, the significant feature of these results is that they illustrate how sharply the extrapolated vaporization curve divides the fluid part of the field into liquid-like and vapour-like regions.

We are grateful to Professor C. Domb for a discussion of this work. A full account of the investigation into the equation of state will be published by one of us at a later date.

Department of Physics,
Queen Mary College,
London.
23rd July 1956.

G. O. Jones.
P. A. Walker.

EUCKEN, A., and HAUCK, F., 1928, *Z. Phys. Chem.*, **134**, 161.
HABGOOD, H. W., and SCHNEIDER, W. G., 1954, *Canad. J. Chem.*, **32**, 164.
HAUSEN, H., 1926, *Z. Tech. Phys.*, **7**, 446.
HOGE, H. J., 1950, *J. Res. Nat. Bur. Stand.*, **44**, 321.
MAYER, J. E., and HARRISON, S. F., 1938, *J. Chem. Phys.*, **6**, 87, 101.
MICHELS, A., and STRIJLAND, J., 1952, *Physica*, **18**, 613.
MICHELS, A., WASSENAAR, T., WOLKERS, G. J., and DAWSON, J., 1956, *Physica*, **22**, 17.
PALL, D. B., and MAASS, O., 1938, *Canad. J. Res.* B, **16**, 449.
RICE, O. K., 1947, *J. Chem. Phys.*, **15**, 314.

Hole Injection at Metal–Semiconductor Point Contact

In a recent paper, Gunn (1954) has discussed the theory of the metal–semiconductor contact, assuming a linear geometry. His theory can easily be extended to include the case of point contact with spherical geometry. This will be done subsequently, n-type semiconductor being assumed throughout.

The point contact is assumed to be a metallic hemisphere of radius a, penetrating into the semiconductor. b is the radius of the boundary between the space charge region and the bulk of the semiconductor, and r is the distance from the centre of the contact (actually $b-a \ll a$). The potential distribution near the metal–semiconductor boundary is of the usual form; ϕ_n and ϕ_p are the quasi Fermi levels for electrons and holes in the semiconductor, ϕ_∞ is the Fermi level at $r=\infty$ (where $\phi_n=\phi_p=\phi_\infty$), and ϕ_m is the Fermi level in the metal. Here $\phi_n=\phi_p=\phi_m$. ϕ_p is constant and equal to ϕ_m over the region $a<r<b$, while ϕ_n is constant and equal to ϕ_∞ over the region $b<r<\infty$ (Shockley 1951).

The equations for the hole and electron current as well as those of continuity and Poisson can be readily written for the case of spherical symmetry. Using these equations and Gunn's assumptions we obtain for the hole concentration in the region $r>b$

$$\frac{d^2(p-p_\infty)}{dr^2}+2r^{-1}\frac{d(p-p_\infty)}{dr}-(p-p_\infty)/L_p{}^2=0 \qquad \ldots\ldots(1)$$

where $L_p=(D_p\tau_p)^{1/2}$ is the diffusion length for holes and the suffix ∞ refers to $r=\infty$. With the transformation $y=r(p-p_\infty)$ equation (1) becomes a linear equation with constant coefficients in y. Using the boundary conditions $p-p_\infty=0$ at $r=\infty$ and $p=p_b$ at $r=b$ we readily obtain the solution for p:

$$p-p_\infty=p_\infty\left(\exp\frac{q\Phi}{kT}-1\right)\left(\frac{r}{b}\right)^{-1}\exp\left[\frac{-(r-b)}{L_p}\right] \qquad \ldots\ldots(2)$$

where Φ is the voltage applied to the contact ($\Phi=\phi_m-\phi_\infty$) and q the electronic charge. j_{pb} the hole current entering the bulk of the semiconductor is given by

$$j_{pb}=qD_p p_\infty\left(\exp\frac{q\Phi}{kT}-1\right)(L_p{}^{-1}+b^{-1}). \qquad \ldots\ldots(3)$$

The flow of electron current has to be considered in the region $a<r<b$. This is a spherical shell, its thickness $b-a$ being very small compared with its radius a. Therefore, the flow of electron current in this region can be assumed to be linear, and may be expressed at $r=b$ as by Gunn (1954).

The injection ratio γ for the extrinsic range can be expressed by $\gamma = \theta/(1+\theta)$ where

$$\theta = \frac{j_{\mathrm{pb}}}{j_{\mathrm{nb}}} = \left(\frac{\kappa \boldsymbol{k} T}{2\pi}\right) q^{-1} \left(\frac{\mu_{\mathrm{p}}}{\mu_{\mathrm{n}}}\right) n_i^{2} (N_{\mathrm{d}} - N_{\mathrm{a}})^{-5/2} (L_{\mathrm{p}}^{-1} + b^{-1}) \exp \frac{q\Psi}{\boldsymbol{k}T} \Big/ F\left[\frac{q}{\boldsymbol{k}T}(\Psi - \Phi)\right]. \qquad \ldots\ldots(4)$$

Here κ is the dielectric constant of the semiconductor, Ψ is the height of the potential barrier on the surface at equilibrium, N_{d} and N_{a} are the donor and acceptor concentrations, F is as defined by Gunn (1954), the remaining symbols having their usual significance.

As can readily be seen from equation (4), a small change in the value of the barrier height Ψ will cause an appreciable change in the value of γ. On the other hand the value of Ψ is insensitive to changes in γ. Therefore, an attempt to calculate Ψ from measurements of γ would fail.

Department of Physics,
The Hebrew University,
Jerusalem, Israel.
D. GERLICH.

20th June 1956; in final form 8th August 1956.

GUNN, J. B., 1954, *Proc. Phys. Soc.* B, **67,** 575.
SHOCKLEY, W., 1951, *Electrons and Holes in Semiconductors* (New York : Van Nostrand).

Time Lags in the Intrinsic Electric Breakdown of Solid Dielectrics

Little is known about the time lags involved in the intrinsic breakdown of solid dielectrics. The information reviewed by Whitehead (1951) suggested they probably are not greatly in excess of 10^{-7} second. Some recent measurements (Kawamura *et al.* 1954) with potassium chloride crystals are consistent with this but in other recent work (Inuishi *et al.* 1952) rather longer time lags were observed. During investigations of the electric strength of ionic crystals when subjected to linearly rising voltage impulses, we were led to suspect appreciable time effects. Consequently, measurements were made of the time lag in the breakdown of various ionic crystals and also in polythene and polystyrene.

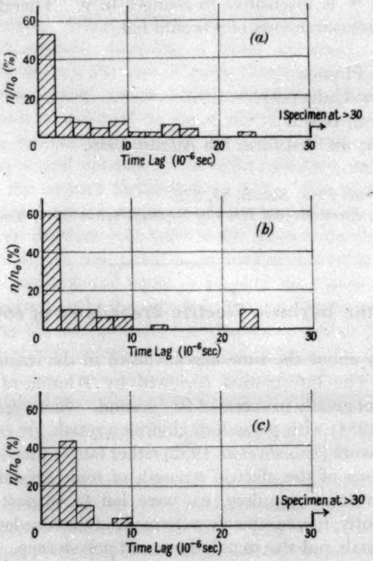
The method of measurement consisted of applying to each specimen a series of 1 : 5000 microsecond impulses until it broke down. The amplitude of each impulse exceeded that of the previous one by about 3% and its waveform was displayed on a calibrated cathode-ray oscillograph and photographed. The voltages at successive instants of time are constant to within 1% for about 50 microseconds from the peak value of a 1 : 5000 microsecond impulse and, since this time is appreciably greater than the longest time lag observed, the voltage applied to the crystal may be considered constant from the peak of the wave until breakdown takes place.

The methods of preparing specimens from ionic crystals and other experimental techniques were similar to those described by Calderwood, Cooper and Wallace (1953). The specimens were annealed and initially free from strains. Recessed specimens of polythene and polystyrene were moulded from sheet material by a method similar to that described by Austen and Pelzer (1946) and Oakes (1948). Graphite electrodes were applied to all of the specimens.

In the case of ionic crystals measurements were made at $-195°$c, $+20°$c and $+220°$c. The results obtained with sodium chloride are illustrated in the figure and the behaviour of potassium chloride and of potassium bromide was sensibly the same. The distribution of time lags at 20°c and at $-195°$c may be represented by the equation

$$n = n_0 \exp(-t/T)$$

where n is the number of specimens breaking down with time lags exceeding t microseconds, n_0 is the total number of specimens tested, T is a constant of magnitude about 1 microsecond.



(*a*) 64 specimens at $+20°$c ; (*b*) 53 specimens at $-195°$c ; (*c*) 34 specimens at $+220°$c.

At 220°c the time lags were grouped about 3 microseconds and the above formula does not represent the distribution. Also at this temperature the collapse of voltage across the specimens occurred over a period of about ¼ microsecond, but at 20°c and $-195°$c this time could not be measured.

Measurements with polythene and polystyrene were made only at room temperature. Appreciable time lags were also observed with these materials and their distribution could be represented by the equation given above provided T was about 0·25 microsecond.

The cause of the time lags is at present not known, but no correlation was observed between breakdown voltage and length of time lag, nor between electric strength and specimen thickness, and it is, therefore, unlikely that the phenomenon can be attributed to spurious discharges at the electrodes or in the ambient. It is believed that breakdown at 220°c was influenced by thermal effects and at this temperature intrinsic breakdown did not occur.

It is evident from the magnitude of the effect illustrated in the figure that the practice sometimes used of measuring intrinsic strength by applying a single linearly rising voltage impulse of a few microseconds duration may yield values of electric strength that are appreciably higher than the intrinsic strength because of the possible voltage overshoot. The existence of the above time lags provides an alternative explanation to some of the impulse effects observed by Alger and von Hippel (1949) with KBr and attributed by them to space charge formation.

The experiments described above were carried out in the Electrical Engineering Laboratories of Liverpool University. We wish to thank Mr. W. G. Oakes of the I.C.I. Research Laboratories, Winnington, for his assistance in arranging the supply of polythene sheet.

Electrical Engineering Laboratories,
University of Manchester.
R. Cooper.
D. T. Grossart.

4th May 1956; in final form 14th August 1956.

Alger, R. S., and von Hippel, A., 1949, *Phys. Rev.*, **76,** 127.
Austen, A. E. W., and Pelzer, H., *J. Instn Elect. Engrs* Pt. I, 1946, **93,** 525.
Calderwood, J. H., Cooper, R., and Wallace, A. A., 1953, *Proc. Instn Elect. Engrs*, Pt. III, A, **100,** 105.
Inuishi, Y., Yamanaka, C., and Suita, T., 1952, *Tech. Reports, Osaka University*, **2,** No. 31.
Kawamura, H., Ohkura, H., and Kikuchi, T., 1954, *J. Phys. Soc., Japan*, **2,** 541.
Oakes, W. G., 1948, *Proc. Instn Elect. Engrs*, Pt. I, **95,** 36.
Whitehead, S., 1951, *Dielectric Breakdown of Solids* (Oxford : University Press), p. 91.

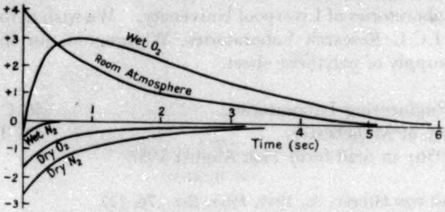
Slow Relaxation Phenomena in Junction Diodes

Slow relaxation effects associated with germanium surfaces have been observed by measurements of surface conductance with transverse electric fields (Low 1955, Morrison 1956, Kingston and McWhorter 1956) of channel conductance in n–p–n structures (Statz, Davis and de Mars 1955) and of surface recombination velocity with transverse fields (Many *et al.* 1956, Nixon and Banbury 1956).

Similar effects have now been found by measurements of the reverse current of a germanium junction diode. The experimental arrangement was as follows: The diode was made by the normal indium alloy process with a base of 0·5 ohm cm n-type germanium. The junction area was 0·2 cm² and the base thickness about 0·020 cm which is much smaller than the diffusion length in the bulk material. A ring-type base connection was made so that the surface opposite to the junction was exposed to the ambient atmosphere. This surface was etched in CP.4. The dimensions of the diode ensured that a substantial proportion of the reverse current originated at the rear surface. The diode was mounted so that each side was hermetically sealed from the other. The junction side was kept in an atmosphere of dry nitrogen to minimize channel effects and on the other side the ambient atmosphere could be varied. Provision was made for illuminating that side through a window.

A battery was connected to apply a reverse voltage to the diode through a resistance and switch. The voltage developed across the resistance by the reverse current was applied to an oscilloscope through a compensating network which balanced out the steady state current. In this way only the variations of current were amplified and displayed.

When a reverse bias voltage was applied it was found that the reverse current took several seconds to reach its final value. The magnitudes and decay times of these effects were dependent on the ambient atmosphere on the rear side. The normal hole storage decay, which has a duration of microseconds, was not observed on this time scale. Some of the oscillograms obtained are illustrated in the figure. None of the decays was found to be exponential. The steady state reverse current was also a function of the ambient atmosphere and variations up to 75% were observed.



Variation of reverse current with time.

These effects could be considerably increased in magnitude when the diode was first illuminated and then the light was interrupted simultaneously with application of the reverse bias voltage. The application of a forward current before switching gave spurious effects due to heating.

The results are qualitatively explicable in terms of the usually accepted model of the germanium surface, with recombination centres at the germanium–germanium oxide interface and slow states associated with the oxide layer. The application of a reverse voltage decreases the carrier concentrations in the base which changes the surface barrier height (Garrett and Brattain 1955). The duration of this change is controlled by the hole storage decay time. There is then a slow variation in the barrier height as the slow states come into equilibrium. The reverse current is governed by the surface recombination velocity which is a function of the barrier height (Stevenson and Keyes 1954) so that the slow changes of barrier height result in variations of reverse current. The increase of the effect with illumination is due to the increase in the change of the carrier concentrations. Variations with ambient may be explained by the dependence of the surface barrier height and the properties of the slow states on the ambient atmosphere (Bardeen and Brattain 1953, Kingston and McWhorter 1956).

The author wishes to thank Dr. K. Hoselitz for his encouragement and advice and Mr. P. Moore for making the measurements.

Mullard Research Laboratories,
Salfords, Redhill, Surrey.
28th September 1956.

T. B. Watkins.

Bardeen, J., and Brattain, W. H., 1953, *Bell Syst. Tech. J.*, **32,** 1.
Garrett, C. G. B., and Brattain, W. H., 1955, *Phys. Rev.* **99,** 376.
Kingston, R. H., and McWhorter, A. L., 1956, *Phys. Rev.*, **103,** 534.
Low, G. G. E., 1955, *Proc. Phys. Soc.*, B, **68,** 10.
Many, A., Morgoninski, Y., Harrick, E., and Alexander, E., 1956, *Phys. Rev.*, **101,** 1433.
Morrison, S. R., 1956, *Phys. Rev.*, **102,** 1297.
Nixon, J. D., and Banbury, P. C., 1956, *Proc. Phys. Soc.* B, **69,** 487.
Statz, H., Davis, L., and de Mars, G. A., 1955, *Phys. Rev.*, **98,** 539.
Stevenson, D. T., and Keyes, R. J., 1954, *Physica*, **20,** 1041.

CORRIGENDUM

On the Theory of Plasma Waves, by F. **Berz.** (Proc. Phys. Soc. B, 1956, **69,** 939.)

In equation (21) the second term on the right-hand side should have an additional factor exp $(-z^2)$.

CONTENTS OF SECTION A

PAGE

Dr. B. H. Bransden, Dr. H. H. Robertson and Dr. P. Swan. The Elastic Scattering of Neutrons by Tritons and ^{3}He 877

Mr. K. L. Kaila and Dr. N. K. Saha. Ionization Loss and the Range–Energy Relation for Low Energy Protons in Metals 888

Dr. P. L. Marsden and Dr. G. A. Doran. The Nucleonic Component in Cosmic Ray Extensive Air Showers 894

Miss P. Rothwell, Mr. B. Wade and Mr. A. Goodings. Directional Measurements on Extensive Air Showers 902

Prof. D. R. Bates and Mr. T. J. M. Boyd. Inelastic Heavy Particle Collisions involving the Crossing of Potential Energy Curves: IV—Ionic Recombination 910

Dr. A. K. Mann and Prof. E. W. Titterton. The Reactions $^{12}C(\gamma, p)^{11}B$, $^{11}B(p, \gamma)^{12}C$ and $^{12}C(\gamma, 3\alpha)$ at 17·63 MeV Excitation 917

Prof. H. Dingle. A Problem in Relativity Theory 925

Prof. W. H. McCrea. A Problem in Relativity Theory: Reply to H. Dingle . 935

Research Notes:

Mr. H. R. Post. Many-Particle Systems: II 936

Dr. R. J. Elliott, Mr. R. D. Lowde and Dr. W. Marshall. On the Multiple Spin Wave Scattering of Neutrons in Ferromagnets 939

Reviews of Books 941

Contents of Section B 944

Subject Index, Section A, Vol. 69 945

Index of Authors (with Titles), Section A, Vol. 69 951

Index to Reviews of Books, Section A, Vol. 69 960

Figure 2. $L = 0{\cdot}065$ cm

Figure 3. $L = 0{\cdot}035$ cm

Figure 4. $L = 0{\cdot}02$ cm

Figure 5. $L = 0{\cdot}035$ cm

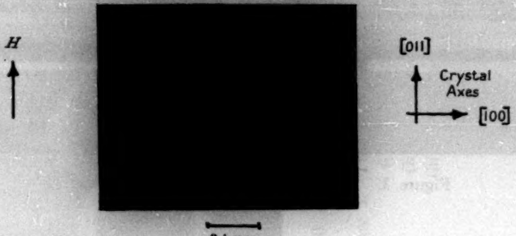


Figure 6. $L = 0{\cdot}035$ cm

Figure 2. Normal lace patterns, $H \simeq 10$ Oe.

Figure 3. ‘Sloping’ lace patterns, $H \simeq 10$ Oe.

Figure 4. Lines of new pattern, $H \simeq 10$ Oe, small normal field.

Figure 5. Chain patterns, $H \simeq 30$ Oe.

Figure 6. Merging of two types of pattern, $H \simeq 30$ Oe.

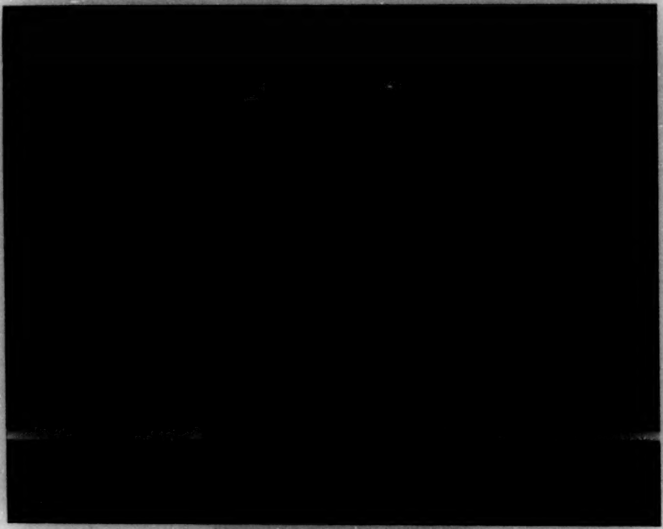


Figure 3. Hall voltage pulse, from sample A at 15°K.

PROC. PHYS. SOC. VOL. 69, PT. 12—B (M. F. CULPIN AND K. W. KEMP)

(i) (ii) (iii)

Figure 3. The effect of drawing upon the x-ray diffraction pattern by transmission.
(i) Undrawn fibre.
(ii) Fibre drawn to a ratio of 3·0.
(iii) Fibre drawn to a ratio of 5·7.

PROCEEDINGS OF THE PHYSICAL SOCIETY
SECTION B, 1956—VOL. 69

SUBJECT INDEX

PAGE
Absorption edge of amorphous selenium, and change with temperature 506
Absoprtion, nuclear magnetic resonance, in polyisobutylene 281
Absorption, optical, in liquid sulphur–selenium co-polymers 1085
Absorption resonance, radio frequency, absence in triethylamine (R) 252
Absorption spectrum of solid arsenic trisulphide (L) 865
Absorption, *see also* Paramagnetic resonance
Absorption, ultrasonic, *see* Ultrasonic
Airy pattern in systems of high angular aperture (L) 854
Alkali halides, electric strength and surface condition (L) 848
Alkali halides, oxygen activated, optical and electrical properties 670
Alkali halides, *see also* Crystals.
Alloy, AuMn, structrual transition and antiferromagnetism (L) 849
Alloys, copper–nickel, saturation magnetization (R) 1050
Alloys, high coercivity, isotropic and anisotropic magnetic viscosity 1189
Alloys, palladium–uranium, magnetic susceptibilities 997
Anisotropy, magnetic, and cold worked texture of titanium 417
Anthracene, scintillation response to 6–30 kev photoelectrons 721
Antiferromagetism and structural transition in alloy AuMn (L) 849
Aperture, high angular, Airy pattern (L) 854
Arcs, gaseous, radiation from (L) 120
Aromatic hydrocarbon crystals photo- and semi-conduction (R) 1162
Arsenic trisulphide, solid, absorption spectrum (L) 865
Autoradiography for study of diffusion (R) 248

Barium titanate ceramics, ageing of dielectric properties 1261
Birefringence from application of intense beam of light to isotropic medium 344
Bismuth telluride, single crystal 606
Bismuth telluride, thermal conducticity 203
Bitter figures, study using electron microscope 647
Breakdown in CdS single crystals, conductivity changes 975
Breakdown, dielectric, *see* Dielectric breakdown.
Breakdown, electric, intrinsic, of solid dielectrics, time lags (L) 1351
Breakdown of gases subject to crossed electric fields 33
Breakdown, vacuum, and pre-breakdown current 651
Bubble, ideal, undamped free pulsations 893

Caesium bromide, thermal expansion (R) 679
Carrier accumulation in germanium 697
Carrier, minority, life-time measurement in Si 791
Cathode potential minimum, ion oscillations in (R) 114
Cobalt, variation of spectral emissivities with temperature 339
Cold worked texture and magnetic anisotropy of titanium 417
Collisions, interelectronic, effect on electronic conductivity (R) 842
Composite media, elastic and thermo-elastic properties 808
Composite media, electrical conductivity 802
Conduction, electronic, in solids with spherically symmetric band structure 553
Conductivity, electrical, on anisotropic thin films 449
Conductivity, high frequency, in semiconductors 528
Conductivity, *see also* Electrical
Conductivity thermal, *see* Thermal conductivity
Conductor, electrostatic centre 175
Contact, point, metal-semiconductor, hole injection at (L) 1350
Contacts, *see* Rectifiers, Semiconductor.
Contrast, variation in electron microscope images with size of objective aperture 643

PAGE

Cooling surface, and convection of heat from wire moving close to it 731
Co-polymers, liquid sulphur–selenium, optical absorption in 1085
Copper films, evaporated, high vacuum x-ray study of structure 180
Cottrell atmosphere, effect of temperature 459
Creep measurements for determination of dynamic moduli and internal friction of high polymers 885
Creep in metals, high-temperature, mechanism with reference to polycrystalline lead 1173
Crystal, form of effective electronic potential 1124
Crystal, NaCl, grown by Kyropoulos method, x-ray study of structure 878
Crystal structure, anomalous interplanar spacings in evaporated deposits of face-centred cubic metals 260
Crystal structure of rare gases (R) 840
Crystallography of martensitic transformations and surface dislocations 1276
Crystals, alkali halide, defects in, and luminescence and electrical conductivity, 396, corr. 591
Crystals, aromatic hydrocarbon, photo- and semi-conduction (R) 1162
Crystals, body-centred cubic, density of electronic states in 619
Crystals, disorientation of single crystal substrate due to epitaxial stress (R) 583
Crystals of hexagonal selenium, epitaxial growth (R) 687
Crystals, monocrystals and intrinsic block structure, temperature dependence of coherent light scattering (R) 255
Crystals, organic, scintillation decay times 358
Crystals, Si–Fe, lozenge and tadpole domain structures in 1109
Crystals, single, of alkali halides, paramagnetic resonance of Mn in (R) 833, 837
Crystals, single, CdS conductivity changes due to breakdown 975
Crystals, single, Cu, lattice spacings of thin electrodeposits of Co and Ni 432
Crystals, single, iron monotelluride, electrical characteristics (L) 860
Crystals, single, scattering of light by 350
Crystals, single, of selenium, bending (L) 771
Crystals, single *see also* Silicon iron, Single crystal.
Crystals, vacuum evaporated, of Cu and Ag, stacking faults (R) 1059
Current gain in formed point contact n-type Ge transistors 712
Current gain at L–H junctions in germanium 705
Current, pre-breakdown, and vacuum breakdown 651

Defects in alkali halide crystals, luminescence and electrical conductivity 396, corr. 591
Deformation, mechanical, of KCl crystals caused by voltage application 1287
Detectors, CuS, anomalous rectification of (R) 958
Diamond, semiconducting, electrical and optical properties 329
Diamonds, type IIb, electrical measurements on (R) 410
Dielectric breakdown in solids 21, corr. 866
Dielectric properties of barium titanate ceramics, ageing 1261
Dielectrics, liquid, space charge phenomena in 61
Dielectrics, solid, time lags in intrinsic electric breakdown (L) 1351
Diffusion of Cr in Ti, comparison of radioactive tracer methods (R) 250
Diffusion of Cr in Ti studied by autoradiographic method (R) 248
Diffusion, radiative transfer with distributed sources 1077
Diodes, junction, slow relaxation phenomena in (L) 1353
Direction finder for recording rapid bearing fluctuations of short radio waves 311
Discharge, gas, *see* Gas discharge.
Discharge mechanism in high-vacuum, cold-cathode pulsed x-ray tube 748
Discharge, positive column, in longitudinal magnetic field 468
Dislocations, continuous distribution : surface dislocations and crystallography of martensitic transformations 1276
Dislocations in silicon, etching technique for revealing (L) 689
Dislocations, supersonic, and in dispersive media 1013
Dislocations, *see also* Cottrell atmosphere.
Dispersive media, dislocations in 1013
Dissociation energies of N_2 and H_2, electron impact studies (R) 682

PAGE
Domain structure, changes with reduction in depth of single crystal of SiFe . . 1200
Domain structure and magnetization curves 497
Domain structures, lozenge and tadpole, in Si–Fe crystals 1109

Editorial 1
Elastic properties of organic glasses in solid–liquid region at ultrasonic frequencies. 1312
Elastic and thermo-elastic properties of composite media 808
Electric fields, crossed, and breakdown of gases. 33
Electric strength, and mechanical deformation of KCl crystals caused by voltage application 1287
Electric strength and surface condition of alkali halides (L). 848
Electrical conductivity of composite media 802
Electrical conductivity of hexagonal selenium 743
Electrical conductivity, *see also under* Conductivity
Electrical measurements on natural galena at low temperatures 796
Electrical resistivity, *see* Resistivity
Electrodeposits, thin, of Co and Ni on Cu single crystal, lattice spacings . . 432
Electromagnetic waves, scattering by conducting spheres and discs . . . 1038
Electromechanical wave source, power radiated by 139
Electron distribution function in insulators at high field strength (R) . . 685
Electron energy losses, characteristic, in solids 1237
Electron impact studies of dissociation energies of N_2 and H_2 (R) . . . 682
Electron–lattice interaction, effect on phonon frequencies 1095
Electron microscope for study of Bitter figures 647
Electron microscopy, stereoscopic reflection (R) 484
Electron microscopy, variation of contrast of image with size of objective aperture . 643
Electron mobility, temperature dependence in Ge (R) 845
Electronic conductivity, influence of interelectronic collisions and of surfaces (R) . 842
Electrostriction, molecular theory 153
Electronic motion in gases and method of free paths (L) 769
Electronic states, density in body-centred cubic crystals 619
Electrons, momentum distribution in solids: II—Application to Li . . . 1249
Electrostatic centre of a conductor 175
Elevation, low angles, radar measurement 321
Epitaxial stress, disorientation of single crystal substrate by (R) 583
Equation of state of solids (R). 838
Etalon, double, phase adjustment. 1309
Etching technique for revealing dislocations in silicon (L) 689
Explosively loaded thin-walled steel cylinders, spatial distribution of fragments . 1089

Ferrites, iron-deficient manganese, angles between magnetic spin directions . . 1319
Ferromagnetic domain nucleation in SiFe 145
Ferromagnetic resonance, anisotropy correction in 993
Ferromagnetism, Heisenberg theory, high temperature expansions in (L) . . 486
Field emission, effect on behaviour of semiconductor contacts (L) 491
Films, anodic aluminium oxide, sealed in boiling water (R) 1055
Films, copper, evaporated, x-ray study in vacuum of structure 180
Films, selenium, amorphous, volume generated currents and secondary effects in . 1139
Films, thin, anisotropic, electrical conductivity 449
Fluids, irreversible processes, molecular theory 269
Fracture, delayed, of sintered alumina 1293
Fracture of glass by spherical indenters 47
Fractures, brittle, near equilibrium 981
Fragments of explosively loaded thin-walled steel cylinders, spatial distribution 1089
Freezing of water droplets 1028
Frequency response of optical systems 562
Friction, internal, in high polymers, from creep measurements 885
Friction, surface, of solids, general theory. 921

PAGE

Galena, natural, electrical measurements at low temperatures 796
Galvanomagnetic effects, anisotropic, in semiconductors 1101
Gas, determination of second virial coefficient (L) 693
Gas discharge, polarization of continuous spectrum in (R) 765
Gases, breakdown, in crossed electric fields 33
Gases, electronic motion in, and method of free paths (L) 769
Gases, rare, crystal structure (R) 840
Germanium, absorption of 39 kMc/s (39 Gc/s) radiation (L) 488
Germanium, carrier accumulation in 697
Germanium contacts, turnover in (R) 763
Germanium, current gain at L–H junctions 705
Germanium, excess noise spectra in (R) 242
Germanium p–n junction diode, measurement of noise spectra 231
Germanium p–n junctions, growth of excess current 539
Germanium, temperature dependence of electron mobility (R) 845
Germanium, time dependent changes of surface lifetime in presence of electric fields (L) 487
Germanium transistors, n-type point contact, current gain 712
Glass, fracture by spherical indenters 47
Glass, indentation hardness as energy scaling law 55
Glasses, organic, elastic properties in solid–liquid region at ultrasonic frequencies 1312
Graphite, vibrational frequency distribution 1326

Halide phosphors, halide activated 934
Hall coefficient, measurement at low temperatures: InSb 1269
Hall effect in semiconductors (L) 862
Heat capacity of silver below 4·2°K (R) 482
Heat, convection from wire moving close to cooling surface 731
Heat flow in region bounded by circular cylinder (R) 256
Helium 3, influence on lambda point (R) 408
Helium, liquid, *see* Liquid helium.
Hole injection at metal–semiconductor point contact (L) 1350

Image, diffraction, aberration-free, phase distribution near focus 823
Impurity atoms, effect of temperature on atmosphere of 459
Impurity centres, divalent, in semiconductors (R) 1335
Impurity distribution in semi-infinite solidified melt (L) 119
Impurity scattering in semiconductors 76
Indentation of glass, fracture by spherical indenters 47
Indentation of glass, hardness as energy scaling law 55
Indium antimonide, low temperature magneto-resistance anomalies (R) 403
Indium antimonide, radiative recombination in (L) 689
Indium antimonide, recombination theory for 661
Insulators, electron distribution function at high field strength (R) 685
Ionization coefficients, precision measurements, in uniform static fields 2
Ionosphere, effective tilts at places 1000 km apart 301
Ionosphere, electron content, measurement by lunar echo method (R) 953
Ions, Hg^+ and A^+, bombardments of elements by 1148
Iris, stabilized retinal image (L) 1165
Iron group, anhydrous chlorides, magnetic properties : I—Magnetic resonance 210
Iron group, anhydrous chlorides, magnetic properties : II—Cryomagnetic anomalies 220
Iron group cyanides, *see also* Potassium ferricyanide.
Iron monotelluride, single crystal, electrical characteristics (L) 860
Iron and silicon–iron, spectroscopic splitting factors (R) 1052
Iron, variation of spectral emissivities with temperature 339
Irreversible processes in fluids, molecular theory 269
Isotropic medium, birefringence resulting from application of intense beam of light 344

PAGE
Lambda point, influence of ^{3}He (R) 408
Lattice frequency, principal, of MgO 129
Lattice spacings, *see* Crystals.
Lead, polycrystalline, and mechanism of high-temperature creep in metals 1173
Lead sulphide group of compounds, theory of photoconductivity in 613
Light, intense beam, and birefringence in isotropic medium 344
Light scattering, coherent, in monocrystals and intrinsic block structure, temperature dependence (R) 255
Light, scattering by single crystals 350
Liquid helium I, flow through narrow channels 1112
Liquid helium 3, specific heat and entropy 1117
Liquid helium, ^{3}He–^{4}He, Bose–Einstein liquid model (R) 1345
Liquid helium, *see also* Helium.
Liquids, pure, ultrasonic absorption in, over frequency range 100–200 Mc/s, 869, corr. 1171
Liquids, supercooled, velocity and absorption of high frequency ultrasonic waves in 965
Liquids, ultrasonic relaxation theory 814
Low temperature susceptibility and magnetic specific heat of K$_3$Fe(CN)$_6$ 1231
Low temperatures, electrical measurements on natural galena 796
Low temperature, specific heats of fluid argon near critical point (L) 1348
Luminescence and electrical conductivity associated with defects in alkali halide crystals 396, corr. 591

Magnetic anisotropy, *see* Anisotropy.
Magnetic double refraction, theory 1133
Magnetic field, longitudinal, positive column in 468
Magnetic field, population changes with (R) 406
Magnetic properties of anhydrous chlorides of the iron group: I—Magnetic resonances 210
Magnetic properties of anhydrous chlorides of the iron group: II—Cryomagnetic anomalies 220
Magnetic properties of MnAs and MnBi (L) 691
Magnetic resonance, nuclear *see* magnetic Nuclear resonance.
Magnetic spin directions in iron-deficient manganese ferrites 1319
Magnetic susceptibility of palladium–uranium alloys 997
Magnetic susceptibility of polymorphic manganese sulphide 1006
Magnetic susceptibility and specific heat of K$_3$Fe (CN)$_6$ at low temperatures 1231
Magnetic viscosity, comparison in isotropic and anisotropic high coercivity alloys 1189
Magnetization curves and domain structure 497
Magnetization, saturation in copper–nickel alloys (R) 1050
Magneto-resistance anomalies, low temperature, in InSb (R) 403
Manganese, MnAs and MnBi, magnetic properties (L) 691
Manganese sulphide, polymorphic, magnetic susceptibility 1006
Melt, solidified, semi-infinite impurity distribution in (L) 119
Metal films, *see* Films.
Metal surface, irregular, work function and patch field 14
Metals, creep in, *see* Creep.
Metals, face-centred cubic, anomalous interplanar spacings in evaporated deposits (R) 260
Meteor trails, radio echoes from: II—Distribution of meteor magnitudes and masses 83
Meteor trails, radio echoes from: IV—Polarization effects 98
Meteor trails, radio echoes, phase changes and resonance effects 1069
Moiré, two-dimensional strain measurement by 373
Moment, electric quadrupole, of ^{235}U and paramagnetic resonance absorption in ^{235}UCl$_3$ (L) 1167
Moon, radio echoes from 901

Nickel, variation of spectral emissivities with temperature 339
Nitrogen, active, presence of atoms in, spectroscopic evidence (R) 956
Noise spectra, excess in Ge (R) 242
Noise spectra of Ge p–n junction diode, measurement 231

PAGE

Nuclear magnetic resonance absorption in polyisobutylene . . . 281
Nucleation, sublimation nuclei . . . 161
Nucleation of supercooled water by various chemicals . . . 165
Nylon fibres, 6·6, orientation of structure . . . 1301

Optical absorption, *see* Absorption.
Optical constants and variations of ellipse of vibration with angle of incidence, tables for quick calculation (R) . . . 955
Optical systems, frequency response . . . 562
Organic solutions, liquid, energy transfer from solvent to solute under u.v. excitation . 593
Oscillations, ion, in cathode potential minimum (R) . . . 114

Paramagnetic resonance absorption in ^{235}UCl$_3$ and nuclear electric quadrupole moment of ^{235}U (L) . . . 1167
Paramagnetic resonance in chromic methylamine alum . . . 633
Paramagnetic resonance in complex cyanides of iron group : I—Experiment . . . 1205
Paramagnetic resonance in complex cyanides of iron group : II—Theory . . . 1216
Paramagnetic resonance of Mn in single crystals of alkali halides (R) . . . 833, 837
Paramagnetic resonance spectrum of Fe^{3+} in cubic field of MgO (L) . . . 1169
Paramagnetic resonance of uranium ions (L) . . . 858
Paramagnetic resonance, *see also* Nuclear magnetic resonance.
Patch field and work function of irregular metal surface . . . 14
Phase adjustment of double etalon . . . 1309
Phase distribution near focus in aberration-free diffraction image . . . 823
Phonon frequencies, effect of electron–lattice interaction . . . 1095
Phosphors, halide activated halide . . . 934
Phosphors, *see also* Alkali halides.
Phosphorus, pentavalent, theoretical study . . . 1297
Photoconduction and semiconduction of aromatic hydrocarbon crystals (R) . . . 1162
Photoconductivity, theory, in lead sulphide group of compounds . . . 613
Photoelectrons, 6–30 kev, scintillation response of anthracene . . . 721
Plasma interaction and conduction in semiconductors (L) . . . 961
Plasma waves, theory . . . 939
Point contacts, *see under* Contacts, Germanium.
Polymers, high, dynamic moduli and internal friction from creep measurements . . . 885
Population changes with magnetic field (R) . . . 406
Potassium ferricyanide, susceptibility and magnetic specific heat at low temperatures . 1231
Potential, electronic, effective, form in a crystal . . . 1124
Powder striations produced by electric sparks . . . 548
Pulsations, undamped free, of ideal bubble . . . 893

Radar measurement of low angles of elevation . . . 321
Radiation, 39 kMc/s (39 Gc/s), absorption in Ge (L) . . . 488
Radiation from gaseous arcs (L) . . . 120
Radiative recombination in indium antimonide (L) . . . 689
Radiative transfer with central sources of non-uniform directional intensity . . . 577
Radiative transfer with distributed sources . . . 1077
Radio echoes from meteor trails, phase changes and resonance effects . . . 1069
Radio echoes from meteor trails, characteristics: II—Distribution of meteor magnitudes and masses . . . 83
Radio echoes from meteor trails, characteristics : IV—Polarization effects . . . 98
Radio echoes from the moon . . . 901
Radio waves, short, direction finder for recording rapid bearing fluctuations . . . 311
Radioactive tracer methods for study of diffusion (R) . . . 250
Rate processes, description by means of relaxation times . . . 293
Recombination, photon-radiative, in PbSe, PbTe and PbS (R) . . . 115
Recombination theory for indium antimonide . . . 661
Rectification, anomalous, of CuS detectors (R) . . . 958
Rectifiers, point contact, H$_2$S treated PbS (R) . . . 245

PAGE
Reflection spectra, diffuse, of semiconductors and measurement of energy gap . . . 70
Refraction, magnetic double, theory . . . 1133
Relaxation phenomena, slow, in junction diodes (L) . . . 1353
Relaxation times, for description of rate processes . . . 293
Relaxation, ultrasonic, and vibrational specific heat of CS_2 . . . 625
Resistivity, electrical, of sodium between 78 and 372°K . . . 441
Resonance, absorption, *see* Absorption.
Resonance, nuclear magnetic, *see* Nuclear magnetic resonance.
Resonance, *see also* Ferromagnetic, Paramagnetic resonance.
Retinal image, stabilized, of iris (L) . . . 1165

Sand, variation of sound velocity with stress (L) . . . 847
Scattering of electromagnetic waves by conducting spheres and discs. . . . 1038
Scattering of light by single crystals . . . 350
Scattering, light *see* Light.
Scintillation decay times of organic crystals . . . 358
Scintillation response of anthracene to 6–30 kev photoelectrons . . . 721
Second virial coefficient of a gas, method for determining (L). . . . 693
Selenium, amorphous, absorption edge, and change with temperature . . . 506
Selenium films, amorphous, volume generated currents and secondary effects in . 1139
Selenium, hexagonal, crystals, epitaxial growth (R) . . . 687
Selenium, hexagonal, electrical conductivity . . . 743
Semiconducting diamond, electrical and optical properties . . . 329
Selenium single crystals, bending (L) . . . 771
Semiconduction and photoconduction of aromatic hydrocarbon crystals (R) . 1162
Semiconductor contacts, effect of field emission on behaviour (L) . . . 491
Semiconductor contacts, surface barriers at . . . 513
Semiconductors, anisotropic galvanomagnetic effects . . . 1101
Semiconductors, avalanche injection . . . 781
Semiconductors, charge carrier fluctuations, statistics . . . 1020
Semiconductors, defects with several trapping levels (R) . . . 1056
Semiconductors, energy gap measurement from diffuse reflection spectra . 70
Semiconductors, GaSb–InSb system, solid solution in (R) . . . 1339
Semiconductors, Hall effect in (L) . . . 862
Semiconductors, impurity scattering in . . . 76
Semiconductors, high frequency conductivity in . . . 528
Semiconductors, mixed, anomalous Lorenz numbers (L) . . . 851
Semiconductors, PbSe, PbTe, PbS, photon-radiative recombination in (R) . 115
Semiconductors, plasma interaction and conduction in (L) . . . 961
Semiconductors, statistics of divalent impurity centres (R) . . . 1335
Semiconductors, surface state parameters from conductance measurements (R) . 1331
Semiconductors, *see also* Contact, Diodes, Germanium, Silicon.
Silicon, etching technique for revealing dislocations (L) . . . 689
Silicon, measurements of minority carrier lifetimes . . . 791
Silicon–iron, single crystal, domain structure changes and reduction in depth . 1200
Silver, heat capacity below 4·2°K (R) . . . 482
Single crystal bismuth telluride . . . 606
Sintered alumina, delayed fracture . . . 1293
Sodium, electrical resistivity between 78 and 372°K . . . 441
Solids, characteristic electron losses in . . . 1237
Solids, dielectric breakdown in. . . . 21, corr. 866
Solids, momentum distribution of electrons : II—Application to Li . . . 1249
Solids, with spherically symmetric band structure, electronic conduction in . 553
Solids, equation of state (R) . . . 838
Solids, surface friction, general theory . . . 921
Solutions, *see* Organic solutions.
Sound pulses of finite amplitude, propagation and reflection . . . 382
Sound velocity, variation with stress in sand (L) . . . 847

PAGE
Space charge phenomena in liquid dielectrics 61
Sparks, electric, powder striations produced by 548
Specific heat, vibrational, of CS_2 and ultrasonic relaxation 625
Specific heats of fluid argon near critical point (L) 1348
Spectral emissivities of iron, nickel and cobalt, variation with temperature 339
Spectroscopic evidence for presence of atoms in active nitrogen (R) 956
Spectroscopic splitting factors for iron and silicon–iron (R) 1052
Spectrum, continuous, in gas discharge, polarization (R) 765
Sputtering, bombardment of elements by Hg^+ and A^+ ions 1148
Stacking faults in vacuum evaporated crystals of Cu and Ag (R) 1059
Static fields, uniform, precision measurements of ionization coefficients 2
Steel and coated steel surfaces, relation between wear and distance run (L) 857
Stereoscopic reflection electron microscopy 484
Strain measurement, two-dimensional, by moiré 373
Sublimation nuclei 161
Surface barriers at semiconductor contacts 513
Surface condition and electric strength of alkali halides (L) 848
Surface effects, on electronic conductivity (R) 842
Surface lifetime, time dependent changes in Ge in presence of electric fields (L) 487
Suspension of particles in liquid, ultrasonic propagation in 193

Temperature expansions, high, in Heisenberg theory of ferromagnetism (L) 486
Thermal conductivity of bismuth telluride 203
Thermal expansion of caesium bromide (R) 679
Thin films, *see* Films, thin
Thin layers, of barium and calcium on tungsten (R) 1343
Titanium, diffusion of chromium in, comparison of radioactive tracer methods for study (R) 250
Titanium, diffusion of chromium in, studied by autoradiographic method (R) 248
Titanium, magnetic anisotropy and cold worked texture 417
Transfer, energy, from solvent to solute in liquid organic solutions under u.v. excitation 593
Transfer, radiative, with distributed sources 1077
Transistor, Ge n-type point contact, current gain 712
Triethylamine, absence of radio frequency resonance absorption (R) 252
Tungsten, mixed monolayers of barium and calcium on (R) 1343
Turnover in germanium contacts (R) 763

Ultrasonic frequencies, elastic properties of organic glasses in solid–liquid region 1312
Ultrasonic relaxation theory for liquids 814
Ultrasonic relaxation and vibrational specific heat of CS_2 625
Ultrasonic waves, absorption in pure liquids over frequency range 100–200 Mc/s 869 corr. 1171
Ultrasonic waves, high-frequency, velocity and absorption measurements in supercooled liquids 965
Ultrasonics, propagation in suspensions of particles in liquid 193
Ultra-violet excitation and energy transfer from solvent to solute in liquid organic solutions 593
Uranium ions, paramagnetic resonance (L) 858

Vibrational frequency distribution of graphite 1326

Wave source, electromechanical, power radiated by 139
Water droplets, freezing 1028
Water, supercooled, nucleation by various chemicals 165
Wear, relation with distance run for steel and coated steel surfaces (L) 857
Work function and patch field of irregular metal surface 14

X-ray tube, high-vacuum, cold-cathode pulsed, discharge mechanism 748

INDEX OF AUTHORS (WITH TITLES)

PAGE

Agarwal, B. K. : The influence of ^{3}He on the lambda point (R) 408
Agarwal, B. K. : On the Bose–Einstein liquid model for ^{3}He–^{4}He solutions (R) 1345
Ahern, S. A., and Sucksmith, W. : Saturation magnetization in copper–nickel alloys (R) 1050
Ainsworth, L. : Single crystal bismuth telluride 606
Anastassiades, M., and Ilias, D. : On the anomalous rectification of cuprous sulphide detectors (R) 958
Andreae, J. H., Heasell, E. L., and Lamb, J. : Ultrasonic relaxation and the vibrational specific heat of carbon disulphide 625
Andreae, J. H., and Lamb, J : Ultrasonic relaxation theory for liquids 814
Arthur, J. B., Gibson, A. F., and Gunn, J. B. : Carrier accumulation in germanium 697
Arthur, J. B., Gibson, A. F., and Gunn, J. B. : Current gain at L–H junctions in germanium 705
Ashton, H. M., and Guggenheim, E. A. : A new method for determining the second virial coefficient of a gas (L) 693
Austin, I. G., and Wolfe, R. : Electrical and optical properties of a semiconducting diamond 329
Ayres, R. U., and Tredgold, R. H. : On the crystal structure of the rare gases (R) 840

Baker, J. M. : Paramagnetic resonance in chromic methylamine alum 633
Baker, J. M., Bleaney, B., and Bowers, K. D. : Paramagnetic resonance in some complex cyanides of the iron group : I—Experimental results 1205
Banbury, P. C., with Nixon, J. D. : Time dependent changes of surface lifetime in germanium in the presence of electric fields (L) 487
Barlow, G. S., and Standley, K. J. : Spectroscopic splitting factors for iron and silicon–iron (R) 1052
Barmby, D. S., with Ewles, J. : Optical and electrical properties of some oxygen activated alkali halides 670
Barrie, R. : Electronic conduction in solids with spherically symmetric band structure 553
Bates, L. F., and Davis, P. F. : ‘ Lozenge ’ and ‘ tadpole ’ domain structures on silicon–iron crystals 1109
Bates, L. F., and Hart, A. : Magnetization curves and domain structure 497
Bates, L. F., and Hart, A. : Changes in domain structure accompanying reduction in depth of a single crystal of silicon–iron 1200
Bates, L. F., and Leach, S. J. : The magnetic susceptibilities of palladium–uranium alloys 997
Bates, L. F., and Martin, D. H. : Ferromagnetic domain nucleation in silicon–iron 145
Benbow, J. J. : The determination of dynamic moduli and internal friction of high polymers from creep measurements 885
Berz, F. : On the theory of plasma waves 939, corr. 1355
Bhagat, S. M. : The specific heat entropy of liquid helium three 117
Bhide, V. G., Das, J. N., and Khandekar, P. V. : Analysis of H_2S treated PbS point contact rectifiers (R) 245
Bickerton, R. J., and von Engel, A. : The positive column in a longitudinal magnetic field 468
Bilby, B. A., with Bullough, R. : Continuous distribution of dislocations : surface dislocations and the crystallography of martensitic transformations 1276
Billam, E. R., and Browne, I. C. : Characteristics of radio echoes from meteor trails : IV—Polarization effects 98
Birks, J. B., and Brooks, F. D. : Scintillation response of anthracene to 6–30 kev photoelectrons 721
Bleaney, B., Hutchison, C. A., Llewellyn, P. M., and Pope, D. F. D. : Paramagnetic resonance absorption in $^{235}UCl_3$, and the nuclear electric quadrupole moment of ^{235}U (L) 1167

PAGE

Bleaney, B., Llewellyn, P. M., and Jones, D. A.: Paramagnetic resonances of uranium ions (L) . . . 858

Bleaney, B., and O'Brien, M. C. M.: Paramagnetic resonance in some complex cyanides of the iron group: II—Theory . . . 1216

Bleaney, B., with Baker, J. M., and Bowers, K., D.: Paramagnetic resonance in some complex cyanides of the iron group: I—Experimental results . . . 1205

Bowers, K. D., with Baker, J. M., and Bleaney, B: Paramagnetic resonance in some complex cyanides of the iron group: I—Experimental results . . . 1205

Bradley, D. E., Halliday, J. S., and Hirst, W.: Stereoscopic reflection electron microscopy (R) . . . 484

Bradshaw, F. J., and Pearson, S.: The electrical resistivity of sodium between 78°K and 372°K . . . 441

Brodie, I., and Jenkins, R. O.: Mixed monolayers of barium and calcium on tungsten (R) . . . 1343

Bromley, R. H.: Two-dimensional strain measurement by moiré . . . 373

Brooks, F. D. with Birks, J. B.: Scintillation response of anthracene to 6–30 kev photoelectrons . . . 721

Broom, R. F., and Rose-Innes, A. C.: A pulse method for measurement of Hall coefficient at low temperatures: some results on indium antimonide . . . 1269

Browne, I. C., with Billam, E. R.: Characteristics of radio echoes from meteor trails: IV—Polarization effects . . . 98

Browne, I. C., Evans, J. V., Hargreaves, J. K., and Murray, W. A. S.: Radio echoes from the moon . . . 901

Browne, I. C., Bullough, K., Evans, S., and Kaiser, T. R.: Characteristics of radio echoes from meteor trails: II—The distribution of meteor magnitudes and masses . . . 83

Buchanan, R. H., with Robinson, R. B.: Undamped free pulsations of an ideal bubble . . . 893

Buckingham, A. D., and Pople, J. A.: A theory of magnetic double refraction . . . 1133

Buckingham, A. D.: Birefringence resulting from the application of an intense beam of light to an isotropic medium . . . 344

Bullough, R., and Bilby, B. A.: Continuous distribution of dislocations: surface dislocations and the crystallography of martensitic transformations . . . 1276

Bullough, K., with Browne, I. C., Evans, S., and Kaiser, T. R.: Characteristics of radio echoes from meteor trails: II—The distribution of meteor magnitudes and masses . . . 83

Burgess, R. E.: The statistics of charge carrier fluctuations in semiconductors . . . 1020

Busby, J., and Richardson, E. G.: The propagation of ultrasonics in suspensions of particles in a liquid . . . 193

Cade, R., and Vickers, D. O.: The electrostatic centre of a conductor . . . 175

Calvert, W. J. R. (the late): Pre-breakdown current and vacuum breakdown . . . 651

Carte, A. E.: The freezing of water droplets . . . 1028

Carter, C.: A theoretical study of pentavalent phosphorus . . . 1297

Carter, W. S., and Stevens, K. W. H.: The magnetic susceptibilities of two polymorphic forms of manganese sulphide . . . 1006

Champeney, D. C.: The flow of liquid helium I through narrow channels . . . 1112

Champion, K. S. W.: Radiation from gaseous arcs . . . 120

Champness, C. H.: The statistics of divalent impurity centres in a semiconductor (R) . . . 1335

Cohen, S. G., and Weinreb, A.: Energy transfer from solvent to solute in liquid organic solutions under ultra-violet excitation . . . 593

Cooper, R., and Grossart, D. T.: Time lags in the intrinsic electric breakdown of solid dielectrics (L) . . . 1351

Cooper, R., and Wallace, A. A.: The influence on electric strength of mechanical deformation caused by application of voltage to potassium chloride crystals . . . 1287

Coulson, C. A., with Sherry, P. B.: The vibrational frequency distribution of graphite: I—Out-of-plane modes of a single layer . . . 1326

PAGE

Craggs, J. D., with Thorburn, R.: Electron impact studies of the dissociation energies of N_2 and H_2 (R) 682

Craik, D. J.: A study of Bitter figures using the electron microscope . . . 647

Crawford, S. M.: Some elastic properties of organic glasses in the transition region from solid to liquid at ultrasonic frequencies 1312

Crompton, R. W., Dutton, J., and Haydon, S. C.: Precision measurements of ionization coefficients in uniform static fields 2

Crosby, P.: Tables for quick calculation of optical constants and variation of the ellipse of vibration with angle of incidence (R) 955

Coupland, J. H.: Variation of contrast in electron microscope images with size of objective aperture 642

Culpin, M. F., and Kemp, K. W.: The orientation of the structure of 6·6 nylon fibres 1301

Daly, N. R., with Emeléus, K. G.: Ion oscillations in a cathode potential minimum (R) 114

Das, J. N., with Bhide, V. G., and Khandekar, P. V.: Analysis of H_2S treated PbS point contact rectifiers (R) 245

Davies, M., and Price, A. H.: Addendum to: Absence of radio frequency resonance absorption in triethylamine, by J. S. Dryden and R. J. Meakins . . . 252

Davies, R. O., and Lamb, J.: On the description of rate processes by means of relaxation times 293

Davis, P. F., with Bates, L. F.: ‘Lozenge’ and ‘tadpole’ domain structures on silicon–iron crystals 1109

Ditchburn, R. W., Fender, D. H., Mayne, S., and Pritchard, R. M.: A stabilized retinal image of the iris (L) 1165

Domb, C., and Sykes, M. F.: On high temperature expansions in the Heisenberg theory of ferromagnetism (L) 486

Doniach, S., with Fröhlich, H.: Plasma interaction and conduction in semiconductors (L) 961

Donovan, B., and March, N. H.: High frequency conductivity in semiconductors 528

Donovan, B., and March, N. H.: Momentum distribution of electrons in solids: II—Some general results with an application to metallic lithium . . 1249

Doyle, W. P.: The absorption spectrum of solid arsenic trisulphide (L). . 865

Drabble, J. R., and Wolfe, R.: Anisotropic galvanomagnetic effects in semiconductors 1101

Dryden, J. S., and Meakins, R. J., with an addendum by Davies, M., and Price, A. H.: Absence of radio frequency resonance absorption in triethylamine (R) . 252

Dutton, J., with Crompton, R. W., and Haydon, S. C.: Precision measurements of ionization coefficients in uniform static fields 2

Dyer, H. B., and Wedepohl, P. T.: Electrical measurements on type IIb diamonds (R) 410

Elder, J. W.: Halide activated halide phosphors 934

Embleton, T. F. W.: The propagation and reflection of sound pulses of finite amplitude 382

Emeléus, K. G., and Daly, N. R.: Ion oscillations in a cathode potential minimum (R) 114

von Engel, A., with Bickerton, R. J.: The positive column in a longitudinal magnetic field 468

Englman, R., and Sondheimer, E. H.: The electrical conductivity of anisotropic thin films 449

Eshelby, J. D.: Supersonic dislocations and dislocations in dispersive media . 1013

Evans, D. M.: The temperature dependence of the mobility of electrons in germanium (R) 845

Evans, J. V.: The measurement of the electron content of the ionosphere by the lunar radio echo method (R) 953

PAGE

Evans, J. V., with Browne, I. C., Hargreaves, J. K., and Murray, W. A. S. : Radio echoes from the moon . . . 901

Evans, S., with Browne, I. C., Bullough, K., and Kaiser, T. R. : Characteristics of radio echoes from meteor trails : II—The distribution of meteor magnitudes and masses . . . 83

Ewles, J., and Barmby, D. S. : Optical and electrical properties of some oxygen activated alkali halides . . . 670

Ewles, J., and Stead, J. C. : Luminescence and electrical conductivity associated with defects in alkali halide crystals . . . 396, corr. 591

Fayers, F. J., with Rymer, T. B., and Hewitt, S. J. : Stacking faults in vacuum evaporated crystals of copper and silver (R) . . . 1059

Feltham, P. : On the mechanism of high-temperature creep in metals with special reference to polycrystalline lead . . . 1173

Fender, D. H., with Ditchburn, R. W., Mayne, S., and Pritchard, R. M.: A stabilized retinal image of the iris (L) . . . 1165

Finlayson, D. M., and Greig, D. : Electrical measurements on natural galena at low temperatures . . . 796

Finlayson, D. M., with Greig, D., Llewellyn, J. P., and Smith, T. : Some electrical characteristics of single crystal iron monotelluride (L) . . . 860

Flynn, P. T. G. : The discharge mechanism in the high-vacuum cold-cathode pulsed x-ray tube . . . 748

Fochs, P. D. : The measurement of the energy gap of semiconductors from their diffuse reflection spectra . . . 70

Forrester, P. A., and Schneider, E. E. : Paramagnetic resonance of manganese in single crystals of alkali halides (R) . . . 833

Fröhlich, H., and Doniach, S.: Plasma interaction and conduction in semiconductors (L) . . . 961

Fröhlich, H., and Paranjape, B. V. : Dielectric breakdown in solids . . . 21

Fröhlich, H., Paranjape, B. V., Kuper, C. G., and Nakajima, S. : The influence of interelectronic collisions and of surfaces on electronic conductivity (R) . . . 842

Fürth, R. : Temperature dependence of coherent light scattering in monocrystals and intrinsic block structure (R) . . . 255

Garrod, R. I. : Anomalous interplanar spacings in evaporated deposits of face-centred cubic metals (R) . . . 260

Gerlich, D. : Hole injection at metal–semiconductor point contact (L) . . . 1350

Gibson, A. F. : The absorption of 39 kMc/s (39 Gc/s) radiation in germanium (L) . 488

Gibson, A. F., with Arthur, J. B., and Gunn, J. B. : Carrier accumulation in germanium . . . 697

Gibson, A. F., with Arthur, J. B., and Gunn, J. B. : Current gain at L–H junctions in germanium . . . 705

Giovanelli, R. G., and Jefferies, J. T. : Radiative transfer with distributed sources . 1077

Goldenberg, H. : Some numerical evaluations of heat flow in the region bounded internally by a circular cylinder (R) . . . 256

Goldsmid, H. J. : The thermal conductivity of bismuth telluride (R) . . . 203

Goodwin, D. W. : Space charge phenomena in liquid dielectrics . . . 61

Goodwin, D. W., and McLean, T. P. : Radiative recombination in InSb (L) . 689

Goswami, A. : Disorientation of single crystal substrate due to epitaxial stress (R) . 583

Green, H. S. : Molecular theory of irreversible processes in fluids . . . 269

Greenhow, J. S., and Neufeld, E. L. : Phase changes and resonance effects in radio echoes from meteor trails . . . 1069

Greig, D., with Finlayson, D. M., Llewellyn, J. P., and Smith, T. : Some electrical characteristics of single crystal iron monotelluride (L) . . . 860

Greig, D., with Finlayson, D. M. : Electrical measurements on natural galena at low temperatures . . . 796

Grossart, D. T., with Cooper, R. : Time lags in the intrinsic electric breakdown of solid dielectrics (L) . . . 1351

PAGE

Guggenheim, E. A., with Ashton, H. M. : A new method for determining the second virial coefficient of a gas (L) 693
Gunn, J. B. : Avalanche injection in semiconductors 781
Gunn, J. B., with Arthur, J. B., and Gibson, A. F. : Carrier accumulation in germanium 697
Gunn, J. B., with Arthur, J. B., and Gibson, A. F. : Current gain at L–H junctions in germanium 705

Hall, G. G. : The form of the effective electronic potential in a crystal 1124
Halliday, J. S., with Bradley, D. E., and Hirst, W. : Stereoscopic reflection electron microscopy (R) 484
Haneman, D. : Current gain in formed point contact n-type germanium transistors 712
Hargreaves, J. K., with Browne, I. C., Evans, J. V., and Murray, W. A. S. : Radio echoes from the moon 901
Harper, P. G. : Effect of the electron–lattice interaction on phonon frequencies . 1095
Harris, N. L., with Strachan, J. F. : Bombardment of various elements by Hg$^+$ and A$^+$ ions 1148
Hart, A., with Bates, L. F. : Magnetization curves and domain structure . 497
Hart, A., with Bates, L. F. : Changes in domain structure accompanying reduction in depth of a single crystal of silicon iron 1200
Haydon, S. C., with Crompton, R. W., and Dutton, J. : Precision measurements of ionization coefficients in uniform static fields 2
Heasell, E. L., and Lamb, J. : The absorption of ultrasonic waves in a number of pure liquids over the frequency range of 100 to 200 Mc/s 869
Heasell, E. L., with Andreae, J. H., and Lamb, J. : Ultrasonic relaxation and the vibrational specific heat of carbon disulphide 625
Hewitt, S. J., with Rymer, T. B., and Fayers, F. J. : Stacking faults in vacuum evaporated crystals of copper and silver (R) 1059
Hey, J. S., and Parsons, S. J. : The radar measurement of low angles of elevation 321
Hey, J. S., Stewart, G. S., Pinson, J. T., and Prince, P. E. V. : The scattering of electromagnetic waves by conducting spheres and discs 1038
Hilsum, C. : The absorption edge of amorphous selenium and its change with temperature 506
Hirst, W., with Bradley, D. E., and Halliday, J. S. : Stereoscopic reflection electron microscopy (R) 484
Hogarth, C. A. : On the measurement of minority carrier lifetimes in silicon . . 791
Hopkins, H. H. : The frequency response of optical systems 562
House, H., with Morant, M. J. : The work function and patch field of an irregular metal surface 14
Humphreys-Owen, S. P. F. : The scattering of light by single crystals . . 350
Hunter, A. N. : Measurements of the velocity and absorption of high-frequency 965 ultrasonic waves in supercooled liquids 965
Hunter, A. N., with Matsukawa, E. : The variation of sound velocity with stress in sand (L) 847
Hutchison, C. A., with Bleaney, B., Llewellyn, P. M., and Pope, D. F. D. : Paramagnetic resonance absorption in ^{235}UCl$_3$, and the nuclear electric quadrupole moment of ^{235}U (L) 1167
Huxley, L. G. H. : Electronic motion in gases and the method of free paths (L) . 769
Hyde, F. J. : Measurement of noise spectra of a germanium p–n junction diode . 231
Hyde, F. J. : Excess noise spectra in germanium (R) 242
Hyman, R. A. : Epitaxial growth of crystals of hexagonal selenium 687
Hyman, R. A. : The electrical conductivity of hexagonal selenium 743
Hyman, R. A. : Bending of single crystals of selenium (L) 773
Hyman, R. A. : Optical absorption in liquid sulphur–selenium co-polymers . . 1085

Ilias, D., with Anastassiades, M. : On the anomalous rectification of cuprous sulphide detectors (R) 958

PAGE

Jefferies, J. T. : Radiative transfer with central sources of non-uniform directional intensity 577
Jefferies, J. T., with Giovanelli, R. G. : Radiative transfer with distributed sources . 1077
Jenkins, R. O., with Brodie, I. : Mixed monolayers of barium and calcium on tungsten (R) 1343
Jones, D. A., and Smith, T. : An x-ray study of the structure on NaCl of crystal grown by the Kyropoulos method 878
Jones, D. A., with Bleaney, B., and Llewellyn, P. M. : Paramagnetic resonances of uranium ions (L) 858
Jones, G. O., and Walker, P. A. : Specific heats of fluid argon near the critical point (L) 1348
Jull, G. W. : An experimental investigation of characteristic electron energy losses in solids 1237

Kaiser, T. R., with Browne, I. C., Evans, S., and Bullough, K. : Characteristics of radio echoes from meteor trails : II—The distribution of meteor magnitudes and masses 83
Keith, H. D. : An x-ray study in high vacuum of the structure of evaporated copper films 180
Kemp, K. W., with Culpin, M. F. : The orientation of the structure of 6·6 nylon fibres 1301
Kerner, E. H. : The electrical conductivity of composite media 802
Kerner, E. H. : The elastic and thermo-elastic properties of composite media . . 808
Kerr, I. S. : The origin of the hydrated oxide formed when anodic aluminium oxide films are ‘ sealed ’ in boiling water (R) 1055
Khandekar, P. V., with Bhide, V. G., and Das, J. N. : Analysis of H_2S treated PbS point contact rectifiers (R) 245
Krishnan, R. S., and Srinivasan, R. : Thermal expansion of caesium bromide (R) . 679
Kuhn, H. G., and Ramsden, S. A. : The phase adjustment of the double etalon . 1309
Kuper, C. G., with Fröhlich, H., Paranjape, B. V., and Nakajima, S. : The influence of interelectronic collisions and of surfaces on electronic conductivity (R) . 842

Lamb, J., with Andreae, J. H. : Ultrasonic relaxation theory for liquids 814
Lamb, J., with Davies, R. O. : On the description of rate processes by means of relaxation times 293
Lamb, J., with Heasell, E. L. : The absorption of ultrasonic waves in a number of pure liquids over the frequency range of 100 to 200 Mc/s 869
Lamb, J., with Heasell, E. L., and Andreae, J. H. : Ultrasonic relaxation and the vibrational specific heat of carbon disulphide 625
Landsberg, P. T. : Note on turnover in germanium contacts (R) 763
Landsberg, P. T. : Defects with several trapping levels in semiconductors (R). 1056
Landsberg, P. T., and Moss, T. S. : Recombination theory for indium antimonide . 661
Lane, P. E., with Prowse, W. A. : The breakdown of gases subject to crossed electric fields 33
Leach, S. J., with Bates, L. F. : The magnetic susceptibilities of palladium–uranium alloys 997
Leech, J. W., and Manuel, A. J. : Magnetic properties of some anhydrous chlorides of the iron group : I—Magnetic resonance 210
Leech, J. W., and Manuel, A. J. : Magnetic properties of some anhydrous chlorides of the iron group : II—Interpretation of the cryomagnetic anomolies in $CoCl_2$, $FeCl_2$, $NiCl_2$, and $CrCl_3$ 220
Lees, D. G., with Woolley, J. C., and Smith, B. A. : Solid solution in the GaSb–InSb system (R) 1339
Linfoot, E. H., and Wolf, E. : Phase distribution near focus in an aberration-free diffraction image 823
Llewellyn, P. M., with Bleaney, B., Hutchison, C. A., and Pope, D. F. D. : Paramagnetic resonance absorption in $^{235}UCl_3$, and the nuclear electric quadrupole moment of ^{235}U (L). 1167

PAGE

Llewellyn, P. M., with Bleaney, B., and Jones, D. A.: Paramagnetic resonances of uranium ions (L) 858

Llewellyn, J. P., with Finlayson, D. M., Greig, D., and Smith, T.: Some electrical characteristics of single crystal iron monotelluride (L) 860

Low, G. G. E.: A method of evaluating surface state parameters from conductance measurement on semiconductors (R) 1331

Low, W.: Paramagnetic resonance spectrum of manganese in single crystals of alkali halides grown from solutions (R) 837

Low, W.: The paramagnetic resonance spectrum of Fe^{3+} in the cubic field of MgO (L) 1169

Lyons, L. E., and Morris, G. C.: Photo- and semi-conduction of aromatic hydrocarbon crystals (R) 1162

McKim, F. R., and Wolf, W. P.: The susceptibility and magnetic specific heat of potassium ferricyanide at low temperatures 1231

McLean, T. P., with Goodwin, D. W.: Radiative recombination in InSb (L) . 689

Mackintosh, I. M.: Low temperature magnetoresistance anomalies in indium antimonide (R) 403

Mackintosh, I. M.: Photon-radiative recombination in PbSe, PbTe and PbS (R) . 115

Mansfield, R.: Impurity scattering in semiconductors 76

Mansfield, R.: The Hall effect in semiconductors (L) 862

Manuel, A. J.: The magnetic properties of MnAs and MnBi (L) 691

Manuel, A. J., with Leech, J. W.: Magnetic properties of some anhydrous chlorides of the iron group: I—Magnetic resonance 210

Manuel, A. J., with Leech, J. W.: Magnetic properties of some anhydrous chlorides of the iron group: II—Interpretation of the cryomagnetic anomolies in $CoCl_2$ $FeCl_2$, $NiCl_2$, and $CrCl_3$ 220

March, N. H., with Donovan, B.: High frequency conductivity in semiconductors . 528

March, N. H., with Donovan, B.: Momentum distribution of electrons in solids: II—Some general results with an application to metallic lithium 1249

Martin, D. H., with Bates, L. F.: Ferromagnetic domain nucleation in silicon iron . 145

Matsukawa, E., and Hunter, A. N.: The variation of sound velocity with stress in sand (L) 847

Mayne, S., with Ditchburn, R. W., Fender, D. H., and Pritchard, R. M.: A stabilized retinal image of the iris (L) 1165

Meakins, R. J., with Dryden, J. S., with an addendum by Davies, M., and Price, A. H.: Absence of radio frequency resonance absorption in triethylamine (R) . 252

Melink, O. W.: The distribution of impurity in a semi-infinite solidified melt (L) . 119

Morant, M. J., and House, H.: The work function and patch field of an irregular metal surface 14

Morris, G. C., with Lyons, L. E.: Photo-and semi-conduction of aromatic hydrocarbon crystals (R) 1162

Morris, D. P., and Preston, R. P.: Structural transition and antiferromagnetism in the alloy AuMn (L) 849

Mortlock, A. J., and Tomlin, D. H.: The diffusion of chromium in titanium studied by an autoradiographic method (R) 248

Mortlock, A. J., and Tomlin, D. H.: A comparison of three radioactive tracer methods for studying the diffusion of chromium in titanium (R) . . . 250

Moss, T. S., with Landsberg, P. T.: Recombination theory for indium antimonide . 661

Mossop, S. C.: Sublimation nuclei 161

Mossop, S. C.: The nucleation of supercooled water by various chemicals . . 165

Murray, W. A. S., with Browne, I. C., Evans, J. V., and Hargreaves, J. K.: Radio echoes from the moon 901

Nakajima, S., with Fröhlich, H., Paranjape, B. V., and Kuper, C. G.: The influence of interelectronic collisions and of surfaces on electronic conductivity (R) . . 842

PAGE

Neufeld, E. L., with Greenhow, J. S. : Phase changes and resonance effects in radio echoes from meteor trails 1069
Newman, R. C. : The lattice spacings of thin electrodeposits of cobalt and nickel on a copper single crystal 432
Nixon, J. D., and Banbury, P. C. : Time dependent changes of surface lifetime in germanium in the presence of electric fields (L) 487
Nutkins, M. A. E. : The density of electronic states in body-centred cubic crystals 619

O'Brien, M. C. M., with Bleaney, B. : Paramagnetic resonance in some complex cyanides of the iron group : II—Theory 1216
O'Dwyer, J. J. : A note on the electron distribution function in insulators at high field strength (R) 685
Osmond, W. P. : Angles between magnetic spin directions in iron-deficient magnesium manganese ferrites 1319

Paranjape, B. V. : Polarization of the continuous spectrum in a gas discharge (R) 765
Paranjape, B. V., with Fröhlich, H. : Dielectric breakdown in solids 21
Paranjape, B. V., with Fröhlich, H., Kuper, C. G., and Nakajima, S. : The influence of interelectronic collisions and of surfaces on electronic conductivity (R) 842
Parsons, S. J., with Hey, J. S. : The radar measurement of low angles of elevation 321
Pearson, S. : Delayed fracture of sintered alumina 1293
Pearson, S., with Bradshaw, F. J. : The electrical resistivity of sodium between 78°K and 372°K 441
Piercy, N. A. V., Richardson, E. G., and Winney, H. F. : On the convection of heat from a wire moving through air close to a cooling surface 731
Piggott, M. R. : The relation between wear and distance run, for steel and coated steel surfaces (L) 857
Pinson, J. T., with Hey, J. S., Stewart, G. S., and Prince, P. E. V. : The scattering of electro-magnetic waves by conducting spheres and discs 1038
Plessner, K. W. : Ageing of the dielectric properties of barium titanate ceramics 1261
Plummer, A. R. F. : Observations on the growth of excess current in germanium p–n junctions 539
Pope, D. F. D., with Bleaney, B., Hutchison, C. A., and Llewellyn, P. M. : Paramagnetic resonance absorption in $^{235}\mathrm{UCl}_3$, and the nuclear electric quadrupole moment of $^{235}\mathrm{U}$ (L) 1167
Pople, J. A., with Buckingham, A. D. : A theory of magnetic double refraction 1133
Powles, J. G. : Nuclear magnetic resonance absorption in polyisobutylene 281
Preston, R. P., with Morris, D. P. : Structural transition and antiferromagnetism in the alloy AuMn (L) 849
Price, A. H., with Davies, M., Addendum to : Absence of radio frequency resonance absorption in triethylamine, by J. S. Dryden and R. J. Meakins 252
Price, P. J. : Anomalous Lorenz numbers in mixed semiconductors (L) 851
Prince, P. E. V., with Hey, J. S., Stewart, G. S., and Pinson, J. T. : The scattering of electromagnetic waves by conducting spheres and discs 1038
Pritchard, R. M., with Ditchburn, R. W., Fender, D. H., and Mayne, S. : A stabilized retinal image of the iris (L) 1165
Prowse, W. A., and Lane, P. E. : The breakdown of gases subject to crossed electric fields 33
Pursey, H. : The power radiated by an electromechanical wave source 139

Ramsden, S. A., with Kuhn, H. G. : The phase adjustment of the double etalon 1309
Rayne, J. A. : Heat capacity of silver below 4·2°K (R) 482
Reekie, J., and Yao, Y. L. : Magnetic anisotropy and cold worked texture of titanium 417
Richards, B., and Wolf, E. : The Airy pattern in systems of high angular aperture (L) 854
Richardson, E. G., with Busby, J. : The propagation of ultrasonics in suspensions of particles in a liquid 193

PAGE

Richardson, E. G., with Piercy, N. A. V., and Winney, H. F.: On the convection of heat from a wire moving through air close to a cooling surface . . . 731
Robinson, R. B., and Buchanan, R. H.: Undamped free pulsations of an ideal bubble 893
Roesler, F. C.: Indentation hardness of glass as an energy scaling law 55
Roesler, F. C.: Brittle fractures near equilibrium 981
Rose-Innes, A. C., with Broom, R. F.: A pulse method for measurement of Hall coefficient at low temperatures: some results on indium antimonide . . 1269
Ross, W. J., with Whale, H. A.: An automatic direction finder for recording rapid fluctuations of the bearing of short radio waves 311
Rubenstein, C.: A general theory of the surface friction of solids 921
Rymer, T. B., Fayers, F. J., and Hewitt, S. J.: Stacking faults in vacuum evaporated crystals of copper and silver (R) 1059

Saksena, B. D., and Viswanathan, S.: Principle lattice frequency of MgO . . 129
Scaife, B. K. P.: On the molecular theory of electrostriction . . . 153
Schneider, E. E., with Forrester, P. A.: Paramagnetic resonance of manganese in single crystals of alkali halides (R) 833
Sharma, R. S.: Equation of state of solids (R) 838
Sherry, P. B., and Coulson, C. A.: The vibrational frequency distribution of graphite: I—Out-of-plane modes of single layer 1326
Singh, Sampooran: Spatial distribution of fragments of explosively loaded thin-walled steel cylinders 1089
Smith, B. A., with Woolley, J. C., and Lees, D. G.: Solid solution in the GaSb–InSb system (R) 1339
Smith, T., with Finlayson, D. M., Greig, D., and Llewellyn, J. P.: Some electrical characteristics of single crystal iron monotelluride (L) . . 860
Smith, T., with Jones, D. A.: An x-ray study of the structure of a sodium chloride crystal grown by the Kyropoulos method 878
Smith, W. A.: The electric strength and surface condition of the alkali halides (L) . 848
Sondheimer, E. H., with Englman, R.: The electrical conductivity of anisotropic thin films 449
Spear, W. E.: Volume generated currents and secondary effects in amorphous selenium films 1139
Spray, C. A. F. T.: A simple etching technique for revealing dislocations in silicon (L) 689
Srinivasan, R., with Krishnan, R. S.: Thermal expansion of caesium bromide (R) 679
Standley, K. J., and Stevens, K. W. H.: The anisotropy correction in ferromagnetic resonance 993
Standley, K. J., with Barlow, G. S.: Spectroscopic splitting factors for iron and silicon-iron (R) 1052
Stead, J. C., with Ewles, J.: Luminescence and electrical conductivity associated with defects in alkali halide crystals . . . . 396, corr. 591
Stevens, K. W. H.: Population changes with magnetic field (R) 406
Stevens, K. W. H., with Carter, W. S.: The magnetic susceptibilities of two polymorphic forms of manganese sulphide 1006
Stevens, K. W. H., with Standley, K. J.: The anisotropy correction in ferromagnetic resonance 993
Stewart, D. T.: Spectroscopic evidence for presence of atoms in active nitrogen (R). 956
Stewart, G. S., with Hey, J. S., Pinson, J. T., and Prince, P. E. V.: The scattering of electromagnetic waves by conducting spheres and discs 1038
Strachan, J. F., and Harris, N. L.: Bombardment of various elements by Hg$^+$ and A$^+$ ions 1148
Stratton, R.: The effect of field emission on the behaviour of semiconductor contacts (L) 491
Stratton, R.: Surface barriers at semiconductor contacts 513

PAGE

Street, R., and Woolley, J. C. : A comparison of magnetic viscosity in isotropic and anisotropic high coercivity alloys 1189
Sucksmith, W., with Ahern, S. A. : Saturation magnetization in copper–nickel alloys (R) 1050
Sykes, M. F., with Domb, C. : On high temperature expansions in the Heisenberg theory of ferromagnetism (L) 485

Thorburn, R., and Craggs, J. D. : Electron impact studies of the dissociation energies of N_2 and H_2 (R) 682
Tillett, J. P. A. : Fracture of glass by spherical indenters 47
Tomlin, D. H., with Mortlock, A. J. : A comparison of three radioactive tracer methods for studying the diffusion of chromium in titanium 250
Tomlin, D. H., with Mortlock, A. J. : The diffusion of chromium in titanium studied by an autoradiographic method 248
Tredgold, R. H., with Ayres, R. U. : On the crystal structure of the rare gases (R) . 840

Vickers, D. O., with Cade, R. : The electrostatic centre of a conductor 175
Viswanathan, S., with Saksena, B. D. : Principle lattice frequency of MgO . . 129

Walker, P. A., with Jones, G. O. : Specific heats of fluid argon near the critical point (L) 1348
Wallace, A. A., with Cooper, R. : The influence on electric strength of mechanical deformation caused by a application of voltage to potassium chloride crystals 1287
Waller, M. D. : On powder striations produced by electric sparks 548
Ward, L. : The variation with temperature of the spectral emissivities of iron, nickel and cobalt 339
Watkins, T. B. : Slow relaxation phenomena in junction diodes (L) 1353
Wedepohl, P. T., with Dyer, H. B. : Electrical measurements on type IIb diamonds 410
Weinreb, A., with Cohen, S. G. : Energy transfer from solvent to solute in liquid organic solutions under ultra-violet excitation 593
Whale, H. A. : Effective tilts of the ionsophere at places at about 1000 km apart . 301
Whale, H. A., and Ross, W. J. : An automatic direction finder for recording rapid fluctuations of the bearing of short radio waves 311
Winney, H. F., with Piercy, N. A. V., and Richardson, E. G. : On the convection of heat from a wire moving through air close to a cooling surface 731
Wolf, E., with Linfoot, E. H. : Phase distribution near focus in an aberration-free diffraction image 823
Wolf, W. P., with McKim, F. R. : The susceptibility and magnetic specific heat of potassium ferricyanide at low temperatures 1231
Wolf, E., with Richards, B. : The Airy pattern in systems of high angular aperture (L) 854
Wolfe, R., with Austin, I. G. : Electrical and optical properties of a semiconducting diamond 329
Wolfe, R., with Drabble, J. R. : Anisotropic galvanomagnetic effects in semiconductors 1101
Wood, C. : Some observations on the theory of photoconductivity in the lead sulphide group of compounds 613
Woods, J. : Changes in conductivity resulting from breakdown in cadmium sulphide single crystals 975
Woolley, J. C., Smith, B. A., and Lees, D. G. : Solid solution, in the GaSb–InSb system (R) 1339
Woolley, J. C., with Street, R. : A comparison of magnetic viscosity in isotropic and anisotropic high coercivity alloys 1189
Wright, G. T. : Scintillation decay times of organic crystals 358

Yao, Y. L., with Reekie, J. : Magnetic anisotropy and cold worked texture of titanium 417

INDEX TO REVIEWS OF BOOKS

PAGE

Allis, W. P., and Herlin, M. A.: *Thermodynamics and Statistical Mechanics* 777
Andrew, E. R.: *Nuclear Magnetic Resonance* 777

Bachman, C. H.: *Physics—A Descriptive Interpretation* 125
Bacq, Z. M., and Alexander, P.: *Fundamentals of Radiobiology* 500
Battelle Memorial Institute (compilers): *Abstracts of the Literature on Semiconducting and Luminescent Materials and their Applications* (1953 issue) 124
Chestnut, H., and Mayer, R. W.: *Servomechanisms and Regulating System Design*, Vol. II 494

Cox, R. T.: *Statistical Mechanics of Irreversible Change* 589
Crawford, A. E.: *Ultrasonic Engineering (with particular reference to high power applications)* 413
Curie, M., and Curie, D.: *Questions actuelles en luminescence cristalline* 963
Curran, S. C.: *Luminescene and the Scintillation Counter* 775

Dévé, C.: *Optical Workshop Principles* 587

Evans, R. M., Hanson, W. T., Jr., and Brewer, W. L.: *Principles of Color Photography* 586

Françon, M.: *Le microscope à contraste de phase et le microscope interférentiel* 590

Garner, W. E.: *Chemistry of the Solid State* 774
Gorter, C. J. (Ed.): *Progress in Low Temperature Physics* 774
Granowski, W. L.: *Der elektrische Strom im Gas*: Band 1—*Allgemeine Probleme der Elektrodynamik der Gase* 1065
Grivet, P., Bernard, M. Y., and Septier, A.: *Optique eletronique*: I—*Lentilles electroniques* 587
Guggenheim, E. A.: *Boltzmann's Distribution Law* 495

Holzbock, W. G.: *Instruments for Measurement and Control* 1171
Hoyle, F.: *Frontiers of Astronomy* 264
Hueter, T. F., and Bolt, R. H.: *Sonics. Techniques for the Use of Sound and Ultrasound in Engineering and Science* 415

Institute of Physics (Peiser, H. S. Rooksby, H. P., and Wilson, A. J. C. (Eds.): *X-ray Diffraction by Polycrystalline Materials* 265
Iwanow, A. P.: *Elektrische Lichtquellen*: *Gasentladungslampen* 778
Jeffreys, H., and Jeffreys, B. S.: *Methods of Mathematical Physics* (3rd Ed.) 1067

Journees de mecanique des fluides (Marseille 1952) 493
Kaiser, T. R. (Ed.): *Meteors* (Proceedings of a Symposium on Meteor Physics, Jodrell Bank Experimental Station, University of Manchester, July 1954) 1067
Kapzow, N. A.: *Elektrische Vorgänge in Gasen und im Vakuum* 1066
Kopal, Z.: *Numerical Analysis* 721
Kuiper, G.: *The Earth as a Planet* 1062

Lawden, D. F.: *Mathematics of Engineering Systems (linear and non-linear)* 123
Lin, C. C.: *The Theory of Hydrodynamic Stability* 695

PAGE

Maréchal, A. : *Imagerie geometrique aberrations* (*Traite d'optique instrumentale*) . 122

Panofsky, W. K. H., and Phillips, M. : *Classical Electricity and Magnetism* . . 585
Pauli, W. (Ed.) : *Niels Bohr and the Development of Physics* 778
Pawsey, J. L., and Bracewell, R. N. : *Radio Astronomy* (*International Monographs on Radio*) : 414
Penman, H. L. : *Humidity* 779
Picht, J. : *Mess- und Prufmethoden der optischen Fertigung*, Vol. 1 122
van der Pol, B., and Bremmer, H. : *Operational Calculus, based on the Two-sided Laplace Integral* (2nd Edn) 493

Sears, F. W., and Zemansky, M. W. : *University Physics* (2nd Edn) . . . 267
Selwyn, E. W. H. : *Theory of Lenses* 266
Smith, K. F. : *Molecular Beams* 585
Sturrock, P. A. : *Static and Dynamic Electron Optics* (*an account of focusing in lens,* deflector and accelerator) 962
Swainger, K. : *Analysis of Deformation* : Vol. II—*Experiment and Applied Theory* . 963
Thomason, J. G. : *Linear Feedback Analysis* 773
Tolansky, S. : *Microstructures of Diamond Surfaces* 589
Topping, J. : *Errors of Observation and their Treatment* 720
Twyman, F. : *Optical Glassworking* 413

Yarwood, J. : *High Vacuum Technique* (3rd Edn, revised) 588